AJS

CVIK - HO 3368 - 82 - POISIO7

CONTENTS

Book Reviews

IN THIS ISSUE

ERVING GOFFMAN was professor of sociology and of anthropology at the University of Pennsylvania when he wrote the paper published in this issue. The paper was accepted before his untimely death. From *The Presentation of Self in Everyday Life* (1956) through *Asylums* (1961), *Stigma* (1963), and *Frame Analysis* (1974), he was always concerned with the social and cultural practices that provide for the understanding of experience. Though he often said that language behavior was too micro-level a phenomenon to attract the interest of sociologists, the present paper and *Forms of Talk* (1981) attest to its interest for him. Speakers' practices, it appears, are no less relevant for understanding their understandings than are the other things they do.

CHRISTOPHER WINSHIP is assistant professor in the Department of Sociology and the Center for Urban Affairs at Northwestern University and a research associate in the Economics Research Center, National Opinion Research Center, University of Chicago. His major research interests are youth unemployment and demographic behavior, particularly as it relates to the transition from youth to adulthood. Currently he is involved with Robert Mare in analysis of the 1964–81 March Current Population Survey and is beginning analysis of the 1940–80 census micro data tapes in an attempt to understand the changes in the behavior of black youths. As part of this effort he is working with models of discrete choice as a method of capturing the interdependency of life-cycle statuses and transitions. His previous research has been in the areas of social networks and the measurement of income inequality and segregation.

ROBERT D. MARE is associate professor of sociology at the University of Wisconsin—Madison and during 1983–84 is a Fellow at the Center for Advanced Study in the Behavioral Sciences. He is engaged in research on changes in the youth labor force and in the transition from school to work since 1940 (with Christopher Winship). In addition, he is continuing his investigations of socioeconomic effects on child mortality and of changes in educational inequality.

STEPHEN COLE is professor of sociology at the State University of New York at Stony Brook. He is currently completing work on several aspects of sociological theory, including the importance of self-selection in the social stratification process, and on the development of public opinion about complex issues such as nuclear power.

MURRAY WEBSTER, JR., is professor of sociology at the University of South Carolina. He recently completed a survey of distributive justice processes for a volume edited by Bachrach and Lawler, and he is beginning a program of theoretical and empirical studies of status cues other than beauty.

JAMES E. DRISKELL, JR., is research psychologist at the U.S. Naval Training Center, Orlando, Florida. Parts of his dissertation research on the functioning of personal characteristics in status generalization were recently published in *Social Psychology Quarterly*. While continuing investigations of status generalization, he is studying interaction effects of behavioral styles (demeanor) as cues to status characteristics.

R. J. HOLTON is currently senior lecturer in sociology at Flinders University of South Australia. He taught previously at the University of Glasgow. He is author of *British Syndicalism* and a number of articles in the field of historical sociology. He is currently working on a study of the transition from feudalism to capitalism.

JERE COHEN is associate professor of sociology at the University of Maryland, Baltimore County. Besides Max Weber and the sociology of religion, his scholarly interests include small groups, adolescence, the sociology of education, and status attainment.

Errata

On page 605 of Carol A. Heimer's review of *Acceptable Risk* by Baruch Fischhoff et al. (*AJS* 88 [November 1982]: 603–5), a parenthesis was omitted from the third sentence of the next to last paragraph. The sentence should read "If ever there were a case to which March and Simon's conclusion (that decision makers satisfice rather than maximize) applied, this is clearly it, but the authors fail to discuss *organizational* limits on either calculative or adaptive rationality." Also, the work "not" was omitted from the first sentence of the final paragraph of the review. That paragraph should read "If a field of acceptable-risk decision making does emerge, one can safely bet that this book will not be the classic work read by everyone in the field. But it is still worth reading now."

Information for Contributors

Editorial Procedures: All papers deemed appropriate for the *AJS* are sent out anonymously to readers. The editors rely heavily on the judgments of those readers, although they are not bound by them. To protect anonymity, only the title should appear on the manuscript. Attach a cover page with the title of the manuscript, your name, and your affiliation. All identifying references and footnotes should appear on a separate page. Manuscripts are accepted for publication subject to nonsubstantive editing with the understanding that the *AJS* has the right of first publication. Submission of a manuscript to the *AJS* is taken to indicate the author's commitment to publish in this *Journal*. No paper known to be under consideration by any other journal will be reviewed by the *AJS*. Upon publication, all rights, including subsidiary rights, are owned by the University of Chicago Press. The author retains the right to use his article without charge in any book of which he is the author or editor after it has appeared in the *Journal*.

Manuscript Acceptance Policy: While it is our policy to require the assignment of copyright on most journal articles and review essays, we do not usually request assignment of copyright for other contributions. Although the copyright to such a contribution may remain with the author, it is understood that, in return for publication, the journal has the non-exclusive right to publish the contribution and the continuing right, without limit, to include the contribution as part of any reprinting of the issue and/or volume of the journal in which the contribution first appeared by any means and in any format, including computer-assisted storage and readout, in which the issue and/or volume may be reproduced by the publisher or by its licensed agencies.

Preparation of Copy

1. Type *all* copy—including indented matter, footnotes, and references—double-spaced, allowing generous margins at top, bottom, and sides of page. Please do not justify right-hand margins. Number all footnotes, including acknowledgment footnote.

2. Type each table on a separate page. Refer to each table in numerical order in the text. Prepare tables without vertical lines.

3. Clarify all symbols with words in the margin of the manuscript. Encircle these and other explanatory notes not intended for printing.

4. Draw figures on white paper with india ink. Original or glossy print of each figure will be required if manuscript is accepted.

5. Include an abstract of not more than 100 words summarizing the findings.

6. *Four copies* of the manuscript and abstract, on white paper, must be submitted. Manuscripts should be addressed to the Editor of the AMERICAN JOURNAL OF SOCIOLOGY, 1130 East 59th Street, The University of Chicago, Chicago, Illinois 60637. *Manuscripts are not returned to authors.* Enclose

a self-addressed stamped postcard, containing the manuscript title, to obtain acknowledgment of manuscript receipt. No other acknowledgment will be sent.

7. Each article or research note submitted to the *AJS* must be accompanied by a check or money order for $10.00, payable to The University of Chicago Press in U.S. currency or its equivalent by postal or express money order or bank draft. Papers will not be reviewed until this fee has been paid. The submission fee will be waived for papers solely by student authors. Submissions from students should be accompanied by a faculty member's certification of their student status. Citizens of countries with restrictions on the export of U.S. dollars may request waiver of the submission fee.

Format of References in Text: All references to monographs, articles, and statistical sources are to be identified at the appropriate point in the text by last name of author, year of publication, and pagination where appropriate, all within parentheses. Footnotes are to be used only for substantive observations. Specify subsequent citations of the same source in the same way as the first one; do not use "ibid.," "op. cit.," or "loc. cit."

1. When author's name is in the text: Duncan (1959). When author's name is not in text: (Gouldner 1963).

2. Pagination follows year of publication: (Lipset 1964, pp. 61–65).

3. For more than three authors, use "et al." For institutional authorship, supply minimum identification from the beginning of the complete citation: (U.S. Bureau of Census 1963, p. 117).

4. With more than one reference to an author in the same year, distinguish them by use of letters (*a, b*) attached to the year of publication: (1965*a*).

5. Enclose a series of references with a single pair of parentheses, separated by semicolons.

Format of References: List all items alphabetically by author (providing the full list of multiple authors) and, within author(s), by year of publication, in an appendix titled "References." Examples follow:

Davis, K. 1963*a*. "Social Demography." Pp. 124–37 in *The Behavioral Sciences Today*, edited by Bernard Berelson. New York: Basic.
——. 1963*b*. "The Theory of Change and Response in Modern Demographic History." *Population Index* 29:345–66.
Goode, W. J. 1967. "The Protection of the Inept." *American Sociological Review* 32:5–19.
Moore, Wilbert E., and Arnold S. Feldman. 1960. *Labor Commitment and Social Change in Developing Areas*. New York: Social Science Research Council.
Sanford, Nevitt, ed. 1962. *The American College*. New York: Wiley.
Weber, Max. (1921) 1968. "Society's Problems." Pp. 12–16 in *Economy and Society*, edited by Guenther Roth and Claus Wittich. New York: Bedminster.

Felicity's Condition[1]

Erving Goffman
University of Pennsylvania

In this paper an effort is made to review the relevance and limitations of
a scattering of near-recent work in sociolinguistics, pragmatics, and con-
versational analysis with respect to a central issue in the sociological study
of social interaction: the taken-for-granted and the inferences made there-
from. The hope is that the line between microsociological studies and so-
ciolinguistics can be shown to be arbitrary, requiring those on each side of
the division to address the concerns of those on the other side.

I

A *presupposition* (or assumption, or implication, or background expec-
tation) can be defined very broadly as a state of affairs we take for granted
in pursuing a course of action. We can perform these acts of faith without
"doing" anything. And even appreciation figures variably. We may never
come to be aware of something our action presupposes; having once been
aware, we may no longer be; having not been aware, we may come to
be; being aware, we may try to conceal this fact from others or to allude
to it indirectly. Yet, according to one reading of the term, if we explicitly
attest to a condition of our action we cease to *presuppose* it, although this
ceasing does not lessen our dependency on it.

By this broad definition, in planning at night to leave at dawn, we
would be presupposing the sun will come up. We do and it will. So what?
We also presuppose that an earthquake will not occur before morning
and drop us forever through a fault, and there is an infinitude of other
possibilities. Clearly, almost all of what we presuppose is footless to any
but those who discuss presuppositions and want to make the point that
there are presuppositions of concern to no one. So it behooves the student
not merely to uncover presuppositions but also to present reasons for

doing so, and not merely the reason just cited. Opportunity abounds. An imaginative analyst ought to be able to show the significance of presuppositions that no one else had ever thought would signify,[2] and certainly every quirk and crisis in social life generates sudden insight in this connection, spreading appreciation that what had been unthinkingly taken for granted should have been given thought. For indeed, we are as unthinking about many of the political and economic conditions of our lives as we are about the sun coming up.

Plainly there are unstated grounds of our action that particular others do not require for their understanding of it and their further response (if any) to it. The contrast is with "social presuppositions," these incorporating a double theme, namely, our tacitly taking something for granted (whether aware of having done so or not), and also unabashedly, even unthinkingly, counting on others involved in the action doing likewise, at least enough so they can easily interpret and understand our action accordingly.[3]

Of social presuppositions I propose to deal with only one sort, those involving language use. After all, more perhaps than any other class of actions, writing or saying makes sense only if the actor intends a meeting of the minds—if only enough to inform self-interestedly. Indeed, almost always in using language we take for granted that what we want to get across will get across (along with the message—true or not—that this is the avowed and controlling purpose of our action). Further, if we could not rely on our listeners grasping the point without extended elaboration, we could hardly afford the time to say anything; similarly, if they could not depend on our taking into consideration what they already know, they could hardly afford the time to listen.

The purpose of this paper, then, will be to consider social presuppositions in language use, particularly speech, not writing. Surely a classical sociological theme. And just as surely I must presuppose, as every student of presuppositions has done and does, the rule that accounts for how I decide which social presuppositions are worth discovering. For as with background expectations in general, there are countless social presuppositions in speech that do not have even academic interest. And the presuppositions that ought to be of interest, themselves involve embarrassment; after all, it should be apparent that spoken (and written) discourse, in context, has the capacity to presuppose anything in the world (including bits of the universe beyond), the sorting of which might well seem to be a hapless, immodest undertaking.

[2] A wonderfully hilarious (and sound) example is provided by Jacques Derrida's 92-page analysis of the presuppositions employed by John Searle in the latter's 10-page reply to Derrida's 25-page critique of speech act theory (Derrida 1977).

[3] In this simplified pragmatic view I follow Stalnaker (1973).

So there are problems. One can take encouragement from the fact that a wide range of social presuppositions in discourse appear to be systematically represented through the ways in which we select, order, and prosodically time and intone our words, thus providing a record that is engagingly objective and sensitive—a workable tracing (indeed, machine-preservable) of practices that could otherwise be rendered only by impressionistic analysis. So one can try to work backward from the verbal consequences of presuppositions to what is presupposed, allowing the direction of analysis to constrain what one examines.

II

There is a philosophical view of presuppositions, variously called semantic, existential, or logical. This view, which is given much weight in linguistics, purports to be concerned only with the truth value of statements (or propositions), not with whatever might be presupposed in the act of making a particular statement in a particular setting—a constraint on concern with language behavior also proposed by the Boy Scouts of America. Accordingly, "entailment" is defined as that relation between two statements such that the entailed one is true if the entailing one is. And (in contrast) one statement can be said to presuppose another if, and only if, the presupposed statement must be true if the presupposing statement is to be true or false in a meaningful way.[4] Take the statement,

> Marsha regretted that John saw the movie last night.

This semantically presupposes (it is felt) that there was a movie last night and that John, who exists, saw it, "regretted" there being a "factive" verb. (It is similarly presupposed that a series of nights have occurred, the last one prior to the temporal point of speaking, but matters are not usually carried that far.) It is suggested that presuppositions of this kind hold even when the polarity of the embedding statement is reversed (as when Marsha does *not* regret that John saw the movie last night), something that cannot be said of entailments.

It may seem that here linguists have become overly enamored of philosophy, but in fact they have reason for their interest in statements of fact in spite of the questionable status of such avowals as guides to what exists in the real world. For as Kiparsky and Kiparsky have shown in a critical article (1970), it seems ". . . that whether the speaker presupposes the truth of a complement contributes in several important ways to de-

---

[4] The consideration of this sense of presupposition is a minor industry in philosophy and linguistics, beginning with Frege and Russell, moving on through Strawson, Black, and Sellars, pausing with Karttunen and the Kiparskys, and now everywhere in full production. For bibliography, see Sag and Prince (1979).

termining the syntactic form in which the complement can appear in the surface structure" (p. 143). If, then, one has interest in the tacit understandings which make social intercourse possible, and if one sees that constraints on sentence structure imposed by the grammar of a language qualify perfectly in this regard, then "semantic" presuppositions will have to figure—sociologically no less than linguistically.

Although many students of language strongly support this perspective on presuppositions (here "implications" might be a better word)—thereby, it seems, tempering any interest in the term apart from their own—others argue that no line can ultimately be drawn between what statements semantically presuppose and what people pragmatically do in stating (see, e.g., Sherzer 1973). Thus, one can widen the semantic approach somewhat to include whatever is relevantly presupposed in the broadly understood, standard meaning of words and clauses on the occasion of uttering them in particular contexts.

Which raises a relevant issue. John, chatting with Marsha, says,

> What did you think of the movie last night?[5]

presupposing with these words that Marsha went to one and to only one film the night before, and so, incidentally, will be able to identify what he is referring to. But Marsha in fact need not actually hold the beliefs about the world (in this particular matter) that John apparently possesses, knowing full well, for example, that she did not actually get out to go to the movie house or, having got to the theater, found it shut. John, informed about his wrong construing of what took place, can feel that Marsha will nonetheless take it that his error was "perfectly understandable," in effect that she can easily follow discourse based on this presupposition. So the import is that Marsha and John do not have to share the same presupposition to allow for smooth discourse; they need merely share an understanding of the presuppositions it would be reasonable to have in the circumstances—a matter not of personal belief but other-oriented assessment.

III

If one shifts from an interest in the truth value of sentences to slightly less rarified concerns, an easier beginning for the study of presuppositions in speech is possible. Let John say to Marsha,

> I went to the movie last night but I didn't like it.

---

[5] An example also employed independently, earlier, and with better effect in a suggestive conference paper given in 1978 by Clark and Marshall (1981).

Apart from the fact that the conjunction "but" implies somewhat se-
mantically that John expects to be entertained by the movies he sees, one
has the "it" to consider. As Bloomfield tells us in an elegant chapter (1933,
chap. 15) now considered optimistic, "it" is a substitute here for something
mentioned in the immediately prior text, namely, "movie"; it is an ana-
phoric expression, an "anaphor" that presumably makes for efficiency in
communication, introducing a short, familiar term for one that might be
less so. This simple possibility—apparently found in some form or other
in all languages—can serve as a start. In English we have the use of
pronouns, articles, and demonstratives with noun phrases, pro-verbs,
adjectival replacement of nouns, the modification of word order, and even
"gapping" and full ellipsis ("zero anaphora").

Substitution terms are a clear-cut example of the fact that what is
actually said need not in itself specify what is meant. This leads us to see
that lurking behind the notion of substitution is the notion of what it is
that is substituted for, that is, a word or phrase that is presumably fully
explicit and readily understandable on its own. To wit: the *antecedent,*
the something that comes before to which the anaphoric expression refers
and from which it acquires its relevant sense. Without the antecedent
expression the anaphoric one would not be fully meaningful, would not
adequately identify what the speaker was talking about. In sum, one can
say (although linguists might not) that an anaphoric expression *presup-
poses* its antecedent, in that without an antecedent insufficient specificity
would result. Perhaps one can also say that the coupling of an anaphoric
expression with an antecedent brings coherence and connectedness across
adjacent statements of a turn at talk.

The task before us now is to expand on these two sides of the reference
equation—substitution terms, on one hand, and antecedents, on the other.

1. First note that the formulation of anaphora I have suggested is crude,
and that each little qualification that is necessary points to another little
way in which we can rely on our hearers seeking out what we have in
mind, and hearers can take it that we assume that they are doing so. For
example, in some circumstances speaker can presuppose that hearers will
accept a pronoun on faith and wait a moment for the "antecedent" that
will provide its reference—what has been called "cataphora" or "back-
ward anaphora," as in,

It was simply lousy that movie I saw.

(Indeed, odd constructions of the "right-dislocation" type are merely an
extreme in this connection; it seems that every uttering of more than a
word or two presupposes that hearers will dutifully exercise a modicum
of cognitive patience and wait for a moment to discover the sense they
are to make out of what they are now hearing.) Further, as Keith Stenning

has argued (1978), every antecedent can be categorized in many different ways, each of which can provide a frame of reference for an anaphor. In truth, the notion of an anaphor "substituting" for its antecedent confuses one possibility with a whole function; for what the antecedent does is to allow the hearer to pick out and identify what it is the speaker is making reference to, and what an antecedent provides is a guide to this determination, not necessarily the identification itself. In consequence, agreement, for example, in number, can be grammatically evaded:

> It was a long movie; I hate ones like that.

2. It should be apparent that there is no reason why the connection between the substitute expression and the antecedent should not be made across adjacent talk of two different speakers, not merely adjacent statements within a single speaker's single turn at talk, the difference being that now the "tying" function of the substitution practice is more pronounced.

When, then, John says,

> What did you think of the movie last night?

Marsha is set up to illustrate the central thrust of a dialogic discourse analysis. For Marsha can construct answers that draw on a variety of practices involving substitution and "deletion":

> I thought it was bad.
> It was bad.
> I thought bad.
> Thought it was bad.
> Bad, I thought.
> Bad.

It is a commonplace of discourse analysis (but nonetheless true) to say that examples in the response set depend for their meaning on immediately prior turn's talk, and, indeed, often would not make much sense without it. Prior turn's talk in effect provides a frame of reference for current turn's talk whose very surface structure will reflect this permissible dependency. Prior turn's talk—as information to have in mind in understanding current turn's—is *presupposed*, at least in the broad sense of that term. John's statement (as Sacks's students, among others, have insistently told us) can establish a conditional relevance on any next act by Marsha, leading both John and Marsha to inspect such next act for what it might imply concerning her feelings about John and his question. Here, in short, one has the sequence implication: the orientation of current turn's talk to prior's, or of next to current, or (shorter still) "contingency." The terms "coherence" and "cohesion" have also figured here, an extension of their

initial reference as to how the text of a single speaker or writer hangs together.[6]

Current sociolinguistics provides us with various directions along which this global notion of contingency can be specified, for example, the child's acquisition of conversational competence as exhibited in its tendency to provide temporally sequenced responses at an earlier age than topically relevant ones (Shugar 1978).[7]

Let me add that this argument about conditional relevance and contingency must be handled carefully. When indeed the immediately prior talk of one speaker is responsively followed by a current statement of another speaker, there remains the issue of whether the interchange projected by first speaker does in fact materialize, or instead (to extend Stenning) reference is made to some other aspect of the circumstances of speaker and hearer which can now be seen to have some relevance. Thus when John says,

> What did you think of the movie last night?

there will be circumstances in which Marsha can *coherently* reply in a way that John might find thematically disappointing.[8] To wit,

> I decided not to go.

Or

> The children were fidgeting so I couldn't concentrate.

Or

> John, can't you stop nattering for a moment?

Or

> I thought you weren't talking to me.

Or

> Stop avoiding the topic.

3. A point is to be made here about the notion of discrete anaphoric terms and discrete antecedents. Reconsider Marsha's first-mentioned set of responses to John's question. Plainly one deals here with something that passes beyond discrete words; one deals with statements taken as a whole. Even in the case of a one-word or one-phrase reply, one could

[6] See esp. Halliday and Hasan (1976). Substitution practices are not, of course, the only means by which utterances across turns can be tied; repetition of a key word is another.

[7] For other relevant work see Garvey (1975), Snow (1977), and McTear (1977).

[8] A wide and useful review of disjunctive answers to questions can be found in Weiser (1975). See also Lakoff (1973b) and McTear (1977).

expand each to display the grammatical properties of a canonical sentence—if one wanted to employ that method of asserting traditional linguistic analysis—in which case it remains that the meaning of whole statements is involved, not merely that of isolated words or phrases. Similarly it can be said that the surface structure of a current statement can presuppose the informational content of the whole of an immediately prior turn's talk. Deep structure ambiguity provides the extreme example.

> Are the chickens ready to eat?

thus depends acutely for its meaning on whether the preceding turn's talk had this kind of structure:

> The asparagus is done.

Or this:

> The cows have been fed.

IV

The notion of the surface structure of statements as an objective record of the consequences of presuppositions must now be refined a little.

1. By the term *sentential utterance* (or utterance for short) I will refer to what could be claimed as an elementary unit of talk, what might also be called a "tone group" or "phonemic clause" or clause spoken without marked pause, or sentence fragment, this last referring to what editing rules regarding ellipsis and error can expand or contract certain (usually short) strings of words into. An encoding unit for thought seems to be involved, a message unit, the minimal package of propositional-like meaning. A *turn's talk*—also considered by some to be an elementary unit of spoken discourse—must contain at least one sentential utterance but may contain more. Utterances neatly nest in turns, the beginning and ending of a turn's talk coinciding with the beginning and ending of an utterance, albeit not necessarily the same utterance. Finally, there is the *move*—some verbal or physical action that is neither an utterance nor a turn at talk, but a strategic position that can be taken up through one or more of either.

2. In recent years, stimulated by the work of Jackendoff, Bolinger, Halliday, Crystal, and others, students have recommended that the prosodic features of an (English) utterance, not merely its syntactic and lexical ones, must be seen to be involved with presuppositions. Typically an utterance, an English one at least, has one point of primary stress or one "tonic nucleus"; typically this occurs late in the utterance, and, whatever else, the stress helps to distinguish decisive or pivotal facts, as speaker

sees them, from ones she assumes are not critical, this serving to help hearer organize what he is hearing.

Returning to our scenario: John says his line and Marsha answers,

I thought it was *bad*.

Presumably Marsha's "I" is already the center of cognitive attention and needs no further help in that regard, any more giving an impression, for example, of undue self-centeredness. On the other hand, if Mary is present, too, and Mary answers first by saying that she thought the movie good, Marsha can properly say,

I thought it was *bad*.

For in this case John's utterance has established a possible contrast between Marsha and Mary as the cognitive environment presupposed for Marsha's reply, and her stress on "I" and "bad" becomes understandable as a help to John and Mary in processing her answer. However, if one starts with John saying, "Did either of you think it was bad?" Marsha can reply (first),

I thought it was bad.

John's utterance has allowed Marsha (and Mary) to assume that the next speaker will not need to make a point of *what* she thinks, only *who* it is that thinks it. Indeed, as already illustrated, heavy elision of the unstressed portion is possible without loss of understandability, allowing Marsha to answer merely with

Me!

along (in this instance) with the application of the English rule that only object pronouns can stand alone, subject pronouns being clitics—the stress placement surviving, however, in spite of the operation of this rule.[9]

One formulation of the organizational role of stress might be mentioned. Sometimes it seems that the unstressed portion of a current utterance substantially repeats what has been offered in prior utterances, especially immediately prior utterance, and that the stressed portion constitutes speaker's addition to this given matter, an addition which, as Halliday (1967, p. 204) says, is presented "as not being recoverable from the preceding discourse." And so one has what has come to be called the *given*

---

[9] In general, prosody critically guides the hearing that is given an utterance, entering into its syntax, its point, and other aspects of its framing. The argument has been made by Gumperz and Kaltman (1980) that the prosodic system of our speech may be as responsive to areal influences as it is to language-locked ones, the implication being that two speakers who acquire the same language in different geographical settings may find themselves elegantly situated to misunderstand each other systematically; while speaking, each will cause the other to make wrong inferences on the basis of wrong presuppositions.

and the *new,* a distinction directly addressed to what can be presupposed by way of recently tabled information, as reflected directly in the prosodic organization of immediately subsequent utterances within or across turn.[10] Understandably, then, the unstressed ("pretonic") portion of the utterance is the portion in which elisions can safely be made.

3. In his paper "Language and Consciousness" (1974), Wallace Chafe suggests an extension of the given/new notion in a social psychological direction. He argues that in discourse what second utterance presupposes in its construction is not so much the immediately prior utterances but, rather, what it is that participants can be assumed to have *at the moment* in their consciousness. It is this assumption that will be marked prosodically, distinguishing by pitch contour and juncture the new from the already present-in-consciousness. As Chafe suggests, "The trick is to arrange the new material so that it will be readily assimilated within the material the addressee's consciousness already contains. The speaker must make assumptions as to what the addressee is conscious of, and transmit his own material accordingly" (1974, p. 112). Whatever else he does, then, Chafe usefully draws attention to the distinction between what is in mind and what is not now but can be recalled thereto, both these domains presumably to be contrasted to what is not known at all. Evidence of this concern for the hearer's consciousness provides impressive documentation of the intimacy and delicacy with which speaker takes hearer into consideration; and at the same time one can begin to see that participants' conceptions of the working of the mind are somehow going to figure significantly in the organization of talk.

The difference between what is merely recallable and what is in consciousness is nicely displayed in speaker's practice of introducing material just in time to serve as a base for the anaphoric bridges that her next turn's utterances will extend:[11]

> MARSHA: Do you remember our old landlord, Mr. Schwartz?
> JOHN: [Nods]
> MARSHA: I saw him on the street yesterday. He looked terrible.

Here it would be correct but miss some of the point to say that Marsha's second two utterances presuppose her first. It might be better to say that the seconds, wanting to get themselves said with relative conciseness and the use of standard formats, insert in immediately prior position the utterance that will allow for this. (Within a single turn at talk, "topical-

---

[10] There has been considerable criticism of the given/new thesis, the argument being that it accounts for only a portion of the cases and that additional formulations are also necessary. See, e.g., Chafe (1976) on contrastive stress and Lehman (1977) on everything. A conceptual refinement of given/new is provided by Prince (1981).

[11] A further treatment is available in Keenan and Schieffelin (1976*b*, esp. p. 338).

ization" and "left-dislocation" constructions seem to accomplish the same thing [Keenan and Schieffelin 1976a].) So there are standard preparatory, ground-laying practices, ways of bringing prerequisite materials to mind, in brief, *locaters*. The pure case is to be found in hardware stores where, routinely, mechanically illiterate householders have to explain what they need without knowing the canonical terms:

> JOHN: Ya know, underneath the sink there's this U-shaped pipe and right at the bottom of the U there's a little sort of screw thing.
> SALESCLERK: Ya.
> JOHN: Well, my little thing has a leak in it.

Locater utterances can be intoned as "try-markers" (Sacks and Schegloff 1979, pp. 18–20), that is, they can be presented tentatively followed by a pause which invites a back-channel signal that recognition has occurred, failing which, a different phrasing may be tried, and so on, until success is achieved—the search jointly appreciated for what it is. There is also the zeroing-in strategy of following one locater with a more specific one, and so on, allowing hearers to come by easy stages to mindfulness of something that they could not otherwise picture. I might add that in sociable small talk between the newly acquainted there may be a search for persons, places, or activities known in common, not merely so that relative social standing can be discovered and network placement made, but also, it almost seems, so that utterances can at last be used that are decently anaphoric. Discovered common interests and common acquaintances supply us with something to allude to—allowing us not so much to say something as to mean something.

4. Next look at "topic," "focus," or "theme." I refer to an organizational feature of talk that does not fit neatly with utterances or even turns. Obviously a topic can be volunteered or proposed in a single utterance; but it can hardly be confirmed into existence until it is taken up in a series of subsequent utterances. (Linguistic usage has it that an utterance in itself can contain a topic—thus allowing for the term "topicalization" as a label for a particular kind of intrasentence construction—but I think this a careless and unfortunate use of terms.) Presumably once a subject-matter theme has been established in a conversation, a current utterance can employ substitution practices expandable not by reference to the immediately prior utterance within or across turns but by reference to the conversation's topic or theme. Indeed, once a theme has been established, current utterance will be obliged (not merely allowed) to exhibit topicality or relevance. And what is found is not merely anaphoric bridges to the topic built by "standard" substitution and elision, but bridges built by what can be read as topical continuity, a somewhat different thing. (The same, as suggested, can be said for another mechanism of coherence,

namely, the repetition in one utterance of a word employed in the prior—
whether within or across turns—as when Marsha answers John's question
by saying, "I liked the movie fine.") Clearly, topic-at-hand participates
in both what is in consciousness and what short-term memory easily
allows to become conscious. In sum, insofar as topic provides a presup-
positional base for anaphoric reference, upcoming speaker is released from
dependence on immediately prior utterance; after all, immediately prior
utterance may be less topical than earlier ones, and thus something whose
claim on what follows can be minimized without necessary loss of co-
herence.

In this reference to the functioning of "topic" as a text-given coherence
link or tie between utterances (whether anaphora is employed or not), I
have followed a common practice in assuming that in itself the term
"topic" is not problematic. But of course it is, and in a manner to widen
greatly the bridge it provides between utterances.

When acquainted persons come together into an informal state of talk
and the greetings have been exchanged, and update questions asked and
answered, a "topic-initial" statement may be made, a "mentionable" or
"reportable" that somehow comes to qualify for elaboration. For example,

> Say, did I tell you, I got a new car last week.

Respondent's uptake of this move can more or less abort the topic effort
with a reply that is rather fully respondent centered:

> I just use my bicycle now.

Or, of course, the reply can confirm topicality with, for example,

> Wow! What kind?

This is a question that gives initial speaker the right-of-way to extend
her theme.[12]

But although respondents can produce a display of relevant uptake
through such replies, they can also get by less dutifully. It turns out that
there is not really any one specific topic to which second speaker (or first
speaker in a next utterance within the same turn) can adhere. As already
suggested in regard to dialogic interchanges, there is always a variety of
aspects of prior talk to hit on as something relative to which to display
continuity. A considerable range of such choices will serve well enough
to display that the current speaker is alive to what has been said and
regardful of it as a source of topic. So only after the fact can anyone
discover what the topic has apparently been all along. Initial speaker can

---

[12] Forms of partial uptake that begrudge the topic-hopeful initiator the story-floor have been
considered much more by students of language acquisition than by students of adult con-
versation; perhaps it is easier to see incivility here in children than in ourselves.

thus find that an unanticipated aspect of what she has said comes to be employed as the reference point for next speaker's utterance. At worst, a "reflexive frame break" can be attempted:

> A: Say, did I tell you, I got a new car last week.
> B: Your voice sounds funny; did you do something to your teeth?

The same can be said for initial speaker's own choice of what it is about her *own* statements that she chooses to follow up with in the same turn:

> A: Say, did I tell you, I got a new car last week? [Suddenly seeing the time] Why am I still talking? I've got to run!

In both cases, it can be argued, the hearer of the apparent change in topic must give weight to the possibility that the speaker has indeed satisfied requirements, has demonstrated aliveness to the topic proposal even if not in the manner anticipated. (Of course, on somewhat the same grounds, individuals whose implied requests for the topic-floor have apparently been denied can attempt to treat the denying uptake as a parenthetical sequence—something to be perfunctorily answered in the manner and voice of an aside—and thereafter revert to their desired theme.) Thus, when a topic *does* appear to have emerged, the utterance which turns out to have started it may itself have been introduced because the prior candidate topic had met with respondent resistance—silences, token turns, halfhearted uptakes—all of which provided the hopeful topic-initiator no replies to build on.[13]

5. Whether or not a clear-cut topic can be identified in a conversation, one is left in doubt concerning the functioning of immediately prior utterance as a reference base for current one. In general it would seem that permissible dependence of a current utterance on one that occurred considerably earlier in the conversation is less pronounced than the dependence possible in regard to immediately prior utterance; yet to a degree everything that gets said early in a conversation can be presupposed in some way by later utterances in it. Thus, a conversation builds up a fund of matters that can be referred to succinctly, providing one of the reasons why we are inclined to "fill in" a latecomer. The problem, then, is that one passes by degree from what can be taken to be in immediate consciousness to what can be more or less readily recalled thereto, the *given* changing gradually to the *recallable*.

Further, when one turns from brief interchanges to, say, longish nar-

---

[13] An implication that can be drawn from the useful paper by Maynard (1980, p. 268 ff.). In general it seems topic development not only allows for respondent's uptake influencing direction but requires it, so that at the extreme, a sort of Brownian movement results, a course quite different from narrative telling, the latter needing only periodic signs that hearer interest has not flagged.

ratives, the locater format becomes somewhat complicated. Instead of laying the groundwork with one utterance ("Ya remember Harry?"), the narrator may feel that a whole informational prologue must be provided before the hearers will be able to properly follow along with the unfolding drama and properly frame described events.[14] And once the story line does begin, out-of-frame information may continue to be interwoven, carried by a self-kibitzing editorial voice. All of this stage-setting material functions as the given for the narrative, but utterances in this prefatory material can themselves—on a lower level—be divided into the new and the given. And, of course, within the narrative frame, characters in the replayed events will provide us through their reported utterances with *embedded* examples of the new and the given, which in fact cannot quite function that way for us listeners; for we, in information state terms, are situated at a higher level. We are situated as listeners to the teller's story, not as listeners to the utterances of characters in the teller's story.

V

Up to now I have assumed that presuppositions help determine the structure of an utterance, and these are to be sought in the text, that is, in what has so far been said in the talk, whether in the immediately prior utterance, the topic, or even less salient mentionings. But plainly, although current utterances might be the proper unit in which to study the *effects* of presuppositions, a part or the whole of the text-to-this-point might not be the proper unit to deal with if one is concerned about *where* the antecedents come from or, indeed, just *what* it is that is presupposed. That an utterance might not always be a good place to look to as a source of antecedents has been suggested by many, including Chafe, and indeed, Chafe's concern regarding consciousness facilitates this shift. He comments that objects and events in the immediate view of the participants (and often properly assumed to be in everyone's consciousness) allow the given to derive not merely from the immediately prior *text* but also from physical events in the immediate situation. Anaphora gives way to "deixis."[15] Further, by our pointing and by other gestural practices, local objects and events not theretofore in the center of attention can be made to be. A shift from unnoticed to in-consciousness is achieved through the directing of sensory attention. And, of course, deictic expressions in the form of demonstrative pronouns and definite articles will be used, these

[14] The informational prologue is described by Labov as the "abstract" and the "orientation" (1972, pp. 363–65).

[15] In practice, linguists do not always sustain the distinction between a text-based and an environmental-based source of antecedents; they sometimes describe as anaphoric *any* substitute term that depends heavily for its reference on something beyond itself.

presupposing a location in time and place, and a speaker-hearer point of view. Moreover, these substitute forms will occur as frequently in the initial utterance of a spate of talk as in following ones, as though, indeed, the physical event at issue were itself the functional equivalent of a first utterance. Thus, the utterance in our domestic scenario,

> What did you think of the movie last night?

employs a pronoun "you" as a substitute for a less handy designation (useful, e.g., when John does not remember Marsha's name), and can do so because the antecedent in question is right there in the flesh, not the text. The use of "I" in Marsha's replies is similarly an object-based, not an utterance-based, substitution. Indeed, when one nominates the "context" to account for the difference between what is said and what is meant, one may forget that in real social life persons come equipped with bits of egocentric evidence as to what they are about and, thus equipped, can *themselves* constitute the context. The utterance,

> Ya alright, Harry?

when spoken by the driver of a newspaper delivery truck as he pauses for the second time that morning before a street vendor means, of course (in Philadelphia), "Do you need any more *Inquirers?*" and Harry immediately knows that is what is meant, because the speaker has driven a context, a framework for interpretation of utterances, right up to Harry's feet.

One kind of object or event that might be of special interest to us is, paradoxically enough, the conversation or encounter taken as a substantive whole—a state of talk officially ratifying a set of participants in jointly sustained communication. Phoning a friend, John offers a minimal greeting and then provides a formulaic display that he is alive to the fact that he has intruded, wants little of the recipient's time, and is in fact prepared to postpone the contact:

> Hi, it's John, ya eating?

Or

> Hello, Love, it's me. You busy?

These succinctly expressed assurances do not draw on prior utterances for the sense that is to be made of them but on the conversation that is imminent, including its likely duration. Further, in placing the call itself, we may have the obligation (and the right) to be the initiator of its termination and so are motivated to produce what Schegloff and Sacks call a pre-closing (1973, pp. 303–9), this move incidentally facilitating the apparent simultaneous closing out of the encounter by both parties.

Here, too, the material that makes sense out of the pre-closing move, allowing its import to be read and its substitutions safely made, is not a prior utterance but the conversation as a whole—a conversation taken as a limited claim on participants' time and not a sequence of meaningful utterances.

Deictic expressions (such as "that one" and "over there") and deictic gestures, such as pointing, are themselves not as innocent or self-evident as they might appear. The underlying assumption seems to be that within arms' reach of speaker and hearer, any object can be directionally referred to or gesturally pointed out (or both), thereby serving as a referential base for substitutive practices. So, too, any object can be distinguished from its like and thereby uniquely identified by means of relative difference with respect to some continuous variable such as hue, height, sound volume, or girth. In all of these cases, statements can be made from the joint perspective of the speaker and hearer because, relative to the referent in question, each participant will easily be able to substitute for the other—a sort of encounter-centric possibility. But, of course, one can rely on this vantage point to locate deictically which star one is talking about as well as which chair, so the notion of proximal must be qualified. The critical matter apparently is that the perceivedly shared location in space of speaker and hearer, and the absence of alternative candidate referents in the field of vision, sound, or some other of our senses, allows for uniquely specified reference.[16] (Something the same can be said of temporal deixis.)

In discussing anaphora, it was suggested that an antecedent "is" not what an anaphor refers to; an antecedent merely helps us discover what the reference is. The same can be said about deixis. Pointing to a car parked at the curb (to extend Geoffrey Nunberg's [1979] argument and example), one can say,

> Well, they didn't steal it. [Our car, left on the street for a long time.]
> He'll get a ticket. [Whoever owns the car will.]
> They can park anywhere. [The police can.]
> That's the new one. [The Mercedes 380 SL, in contrast to the 450.]
> She might be hiding under there. [Our runaway cat.]
> That's what's nice about the suburbs. [Easy parking.]

Finally, observe that anaphoric and deictic reference commonly merge, so that in actual cases it becomes difficult to be sure just what basis of

---

[16] "Shared location" is itself a gloss. "You" refers to addressee, "I" to speaker, both from *speaker's* point of view; "on your right" refers ("your" notwithstanding) to an object from *recipient's* point of view. Only with, say, "under the table" is point of view fully shared by speaker and hearer. Yet deixis is equally involved in all these cases. So the issue is not that location is shared but that speaker can (and hearer can assume speaker can, etc.) construe hearer's location sufficiently to take this point of view into consideration in specifying reference.

reference is at work. When one cardiologist moves over to the screen and says to a colleague,

> Take a look at this film; you won't believe it!

"it" is a substitution that is simultaneously (and redundantly) anaphoric *and* deictic. Indeed, as Tasmowsky-De Ryck and Verluyten (1981) suggest, demonstrative pronouns used in reference to actual objects in the immediate setting can be responsive to the linguistic form associated with the unstated verbal representation of these objects, as when (in English) we use "those" while pointing to a pair of pants or scissors, or (in French) use gender-specific pronouns for present objects which are sexless except for the name that would otherwise be applied to them.

VI

From a consideration of anaphora we have moved on to a consideration of deixis—a shift from what has been said in the setting to what has been seen, heard, or smelled. But since the issue now seems to be what it is that is in mind or can be readily brought thereto, one is perforce required to consider another major factor: namely, what each speaker presupposes his listeners knew about the world and its working before the conversational forgathering itself occurred, and, given their knowledge, that they will be able efficiently to bring it to mind.

It should be evident that, although an anaphoric or deictic basis for cryptic reference can rightly be described as "contextual," this is not quite the case in regard to the basis for presuppositions now under consideration. Assumptions regarding knowledge prior to an occasion are situational in that they depend on who in particular is present in the encounter under consideration; however, *any* occasion when the same two persons come into communicative contact will allow them to make some of the same presuppositions about information.

As considered thus far, the notion of imported information as a basis for laconic reference is somewhat simplistic. Given that John and Mary have a "normal" command of the syntax and vocabulary of English, Marsha can talk to them about anything that she knows about, providing only that she is willing to be sufficiently patient in building up explanatory detail. What one really seems to have in mind in regard to Marsha's presuppositions regarding what John knows is this: How laconic can she be and still be assured he will recognize what it is she has in mind? For example, when John and Marsha are alone together they can refer to a set of persons by first name, nickname, or last name, and this on the first occasion of referring to any one of these persons in the conversation, the assumption being that both John and Marsha will know which of the

17

possible persons by that name is being referred to. The issue, of course, is not merely joint familiarity and intimacy; a shared close friend with a popular first name may require greater specification than a mere acquaintance with an uncommon one.

When talking to *Mary*, not Marsha, John will be able to effectively identify another set of persons by first name, this set only partly overlapping, if at all, with the prior; and yet, through less laconic modes of identification, John will be able to specify to Marsha all the individuals whom he identifies by first name to Mary.

It seems adequate enough to use the phrases "common understanding" and "shared presuppositions" when the effective minimal terms John uses with Marsha are the same ones he would employ with many others; but to the degree that these terms are adequately decodable only by Marsha, something stronger need be said in addition. For in many of the conversations John has with others (whether with familiars or unfamiliars), he will not merely fall back on what can *generally* be taken to be shared, nor, indeed, will he merely tailor his references to the particular mind of his particular hearer; rather he will be inclined to select from whatever he shares with this other, *just* those topics that allow him to employ allusive phrases that *only* the recipient would immediately understand. Thus, his talk will not so much depend on common understanding as seek it out and then celebrate it. Indeed, this gives to ordinary verbal contacts a greater degree of exclusivity and mutual dovetailing than one might otherwise expect—not only "recipient design" (Sacks, Schegloff, and Jefferson 1974, p. 727), then, but the celebration of what it implies. Thus the paradox that a very inconsiderate verbal response to someone can nonetheless be very considered—very "personal"—in the selection and laconic phrasing of allusions, imputations, and names.

With these reservations, then, let us see what can be said about imported information.

The first kind of imported information to consider is what apparently can be assuredly taken for granted, namely, what John knows Marsha knows because he has his own assurances that she has acquired that knowledge, having been witness to her witness. John, seeing Marsha at a movie, can thereafter in talking to her presuppose that she has seen it.[17] Marsha will be able to recall the movie and John knows she will be able to. Plainly the more experience John and Marsha have had together, the larger the corpus of known knowings. (In a sense, what we mean by "together" is the acquisition of experience witnessed by all parties as

[17] Obvious qualifications are here neglected but will figure in the discussion below. John seeing Marsha at the movie may (rightly or wrongly) feel that Marsha (1) has not seen him, (2) may not have seen him, (3) has seen him, whether avowedly (say, through a greeting) or not. On occasion each of these possibilities has distinctive relevance.

18

having been mutually witnessed, thus a mutual knowledge of shared information.) One must add, of course, that through talk, John can convey verbal information to Marsha and thereafter in conversations with her will know she knows it, this being a second route to the development of knowingly shared information.

One can argue, then, that there are various grounds for John having personal assurances that Marsha has a particular bit of information (or the capacity to recollect it), which information she can be assumed to import into conversations, allowing John to there take this information for granted in phrasing utterances. This leads one to consider a second class of imported information that can be taken for granted, this time, however, with personal assurances somehow missing, and varying degrees of confidence put in their place for want of specifically knowing about the other's knowing. Significant here is the range of information an adult can assume other adults possess merely by virtue of being competent members of their age class. On inspection, a great deal of this probabilistically warranted information turns out to be deictically based, radiating from the addressee's spatial and temporal position. On the spatial side there are matters pertaining to geography, community, and state—the sort of thing a resident and citizen is likely to know regarding "local matters." On the temporal side there is orientation for the time and date; for events in the past, the more "central" the event is recognized to be, the longer its claim on memory. In addition, a great deal of this information can be glossed as knowledge of the social structure: the claims of parents over their children; the arrangement between the sexes; the prerogatives of age; the prestige ranking of occupations, religions, ethnicities, and races; the correlates of social class; and the like.

The notion of information that can be taken for granted on probabilistic grounds has some use but speaks to a class of facts that is so immense as not to have much value. The literature, however, provides some specifying suggestions and hints, all speaking to the possibility that knowledge about the world might itself have some organizational features, these themselves providing a more specific basis for warranting informational presuppositions. I provide a scattering:

1. There is the argument presented by Harvey Sacks (1972a) concerning "membership categorization devices" and the role they play in providing a context for identificatory terms.

2. There is the suggestion by Clark and Marshall (1981) regarding the packaging of knowledge. In their example Marsha says,

> I bought a candle yesterday.

She is then in a position to say,

> It was expensive.

Where "it" is uniquely specifying because of the linguistic "co-presence" (to use Clark and Marshall's phrase) of candle. But she can also say,

> And you know, of all things the wick turned out to be missing.

Here she assumes, and correctly, that those who know about candles are likely also to have a received version of a candle's constituent parts, of which "wick" is counted as one. Prince makes a similar point with the example "I got on *a bus* and *the driver* was drunk" (1979, p. 269). People who know about buses can be assumed also to know that they have drivers.

3. A related recommendation is provided by the Schank and Abelson (1977) tradition of work associated with the term "script." If the activities we engage in acquire recoverable images involving standard sequences of phases, then familiarity with an activity readies us to understand cryptic references to such phases, especially those immediately adjacent in sequence performance to the one under consideration (Nelson and Gruende 1979).

4. The notion of "course of action" provides an allied way of examining discourse coherence. Our utterances, it could be argued, can provide evidence of our plan, intent, or course of action (as can our glances and gestures), and if addressees have warrant for entering into such plans, they can orient themselves to circumstances wider than those the utterance itself might seem to establish "literally." Thus (following Clark [1979], p. 4; see also Merritt [1976] and Lakoff [1973a]), when we phone a restaurant and ask if they take Diner's Club cards, the restaurateur can assume we want to come in the near future to eat and that cash is not convenient. He can grasp our intent and some of the circumstances in which we must realize it. Thus, he can meaningfully reply,

> No, we don't take credit cards [thus forestalling a needless, possible second round regarding other credit cards] but we take personal checks.

Or

> Yes, but we're closing tonight for a two-week vacation.

To speak here in speech act terms of a response directed to the "illocutionary force" not "locutionary content" of prior utterance is not quite enough; we may not even know what it is about our implied course of action that our respondent knows needs addressing.

5. I think it can be added that just as the design of an object or the phases of an action (as conventionally understood) can provide a map of what it is an individual is likely to know about, so information itself seems to have organizational features, such that an individual who demonstrates knowledge of one item is likely to know others at that level of

specialization and cultivation. As Schegloff (1972) argued (and very well), occupational, professional, trade, and sports argots allow initiates to calibrate the technical sophistication of a client, colleague, or acquaintance from a few uttered phrases and thereafter pitch their own talk so as to draw fully on the other's familiarity with the scene, while not overtaxing the information which this familiarity implies. Command of a language also has "latent structure" effect; our ability to make, or to demonstrate understanding of, one statement allows others to presuppose with warrant that we will be competent with respect to a wide range of other statements.

6. Roger Schank (1977) has made the interesting suggestion that the kinds of topic-confirming questions that can follow a potential topic-initial utterance might themselves be limited and derive from shared understandings concerning "conversational association categories" such that any choice from within the appropriate list will be heard as appropriate by virtue of being topic supportive. Given a command of the relevant set of categories (where did you get it, how much did it cost, what sort of quality, etc.), a respondent can mechanically grind out an impression of being fully alive to what is going on conversationally; in short, discourse particles (such as "Really?"), and a somewhat standard battery of "WH" (what, why, where, etc.) and yes/no questions can be employed as a resource for follow-up responses perceived as relevant by the recipient.

VII

I have argued that utterances can presuppose not only a prior text and matters available in the "immediate" setting but also imported knowledge, and the case can be made that this last has some organization. However, to say that these three basic cognitive resources can provide frames of reference for discrete substitution terms, the surface form of entire laconic utterances, and prosodic markers does not carry us far enough. Although laconic expressions provide the obvious stimulus for one's looking to "shared" understandings of speaker and hearer as a means of decipherment, such expressions are actually not special in this regard. Cryptic expressions—sentence fragments—are obviously dependent for their sense on something that went before or can be assumed; but their word or phrase antecedents are too, merely less obviously so. And when a well-formed, fully grammatical sentence *is* uttered, one that is meaningful even when excerpted from the face-to-face encounter in which it was spoken, one still can have the problem of shifting from what was said to what in fact was meant in the particular scenario. In brief, when there *is* an easy, literal, relatively context-free reading of a particular utterance, this need not be the reading that was intended and received. In general, then, as Garfinkel (1967, pp. 35–75) noted more than a decade ago, any

expansion of a substitute term would itself require for understanding something unstated that was taken for granted, the antecedent always one step away from the dutiful explicator.

The most evident example perhaps comes from the fact that lexical items can have more than one primary dictionary meaning, and, as suggested, in some cases different meanings of the same word can be intended for the same slot in otherwise identical utterances, the business at hand serving to disambiguate, and not the deep structure of the utterance. "Bring me a pump" refers to one sort of thing when a cyclist calls to his buddy and to something quite different when a shoe salesman says it to a stock clerk. "Who opened the restaurant" (as Sag and Liberman suggest [1975, p. 490]) means one thing to residents in the restaurant's neighborhood and another to its owner when asking an employee why the door was opened to business an hour early.

Another example is found in those cases where hearer's knowledge of the world is expected to override purely textual interpretive procedures such as referring a pronoun back to the nearest noun phrase in the same syntactic position. Thus, Marsha can say unambiguously to John,

> Jimmy Connors came into my brother's shop yesterday and he showed him how to hold his racquet.

But she can also say unambiguously,

> Jimmy Connors came into my brother's shop yesterday and he got his autograph.[18]

John's understanding of the world here takes precedence over usual tying practices.

The same point can be made about the use of conjunctions to link two utterances. Often each of the conjoined utterances will have an explicit literal meaning that makes sense in its own right but no sense when joined to the other, unless, that is, the hearer can follow what is alluded to in one or both of the utterances.[19] Thus a mother says to a child who wants to go outside,

> It's mucky out there and I've enough work to do.

In which case, as Lakoff (1971, p. 119) suggests, the child must make presuppositions, "and the presuppositions he makes, though they have no overt form in the superficial structure of the sentence, very definitely affect the acceptability of the sentence." (The reader will note that here I, following Sacks, presuppose that the person looking after a child will

[18] Such pairs of utterances are discussed by Coulter (1979, pp. 171–73), from which I draw. As Coulter suggests, on occasion beliefs about the world may not disambiguate.

[19] Here I draw from the helpful paper by Lakoff (1971).

be its mother, and that Lakoff presupposes that the model subject is properly a male, not a female.)

The upshot, then, is that when a current utterance does provide a necessary verbal context for some truncated next utterance, current utterance may not only accomplish this through its "literal" or dictionary or "locutionary" meaning, but also not accomplish it through any figurative reading. When John says,

> What did you think of the movie last night?

only John and Marsha may know that neither left the house last night and that a TV movie is at issue—let alone which of that night's movies. John's opening remark provides a verbal context for rendering fragmentary responses unambiguous, such as Marsha's:

> Picture wasn't clear.

But there need be nothing either in the interchange itself or its seeable or hearable context to establish that the transmission of the movie is in question, not its plot. Coherence depends on the interactants having shared last night's events, and this resource may never be made palpable through what the interactants literally say because of it. Thus (to take a further example), John can presuppose that he and Marsha appreciate that all he knows about Marsha's evening was that she was planning to go to *a* movie. In that case, when he says,

> What did you think of the movie last night?

he can assume that Marsha will be sure to tell him its name, probably before she tells him what she thought of it, for he can take it that she will know he must mean to query her about the "whichever" movie she saw, or perhaps about the evening's experience as a whole, the movie as an "outing."

One must also consider interchanges in which a first utterance alludes with careful distance to a hinted meaning (one that is distinguished by the fact that it can be denied, if necessary), and the uptake openly replies thereto:[20]

> JOHN: What did you think of the movie last night?
> MARSHA: I'm sorry, I forgot to thank you. I had a simply wonderful time.

A conclusion, then. Certainly discourse provides an objective tracing of presuppositions. And certainly discourse can be taped, the occasion of its production filmed, and the whole result subjected to repeated close

---

[20] The remedial character of these hints is considered in Goffman (1971, esp. pp. 168–69).

examination. But the record itself will not always be enough. The bi-, ography and prior experience of the participants may have to figure. In the matter of presuppositions, then, that which language allows us to study takes us beyond language to social arrangements that are essentially nonlinguistic. In all of this it is plain that no line can be drawn between those presuppositions (in the broad sense of that term) which figure in traditional linguistics and those which members of neighboring disciplines would study. In all cases culturally specific norms of interpretation, not merely the system-features of any particular language, are involved.

Thus, when anaphoric or deictic reference does occur in the natural state, each is likely not only to overlay the other but also to be inextricably mingled with, and dependent on, the reference resources just considered that are not analyzed by linguists. After all, the matter to which a deictic expression or gesture refers must be one that the hearer can segment and perceive as the speaker does, and such matters do not cover "everything" that is nearby, only matters both participants are trained and ready to perceive. Thus, one of our cardiologists can mention or point to a shading in an X-ray that will suffice to ensure that the other will know what the first is talking about, but the laity could not draw on the immediate scene in this way as a presuppositional base for deictic terms or gestures. They might look in the right direction but would not see. Nor is the arcane perceptiveness of cardiologists needed as an example. When Marsha greets John by saying:

Do you like it?

"it" (in my story) is clearly a deictic pronoun referring to the head of hair that is right there for all to see. But "it" can so refer because Marsha knows that John knows, or at least will accept the reasonableness of his knowing, what it looked like before she had it cut; and furthermore she can assume from their relationship that he has (or should have) some interest in the way she looks, some memory of how she looked, and can be queried on the matter with propriety, even if not always with success. What is at issue is not a physical thing but a contrastive relationship, only one end of which is deictically available. Nor can one simply say that "it" is a substitute for "how my hair was and how it is now," for Marsha is not referring to her hair in general, whatever that might be, but to its cut (or color, or wave, or sheen, or combination thereof), and what she and John mean by *these* terms takes us back to broad cultural understandings they both share with lots of people Marsha knows could never have known how her hair looked (and therefore could not reasonably be asked if they liked it, in the sense of liking it *now*), or having known, have no reason to remember, or, perhaps remembering, could not with modesty on Marsha's part be queried thereon. Nor is propriety

24

an issue only in regard to social relationships. Nunberg can hold up a
pen in a phonetics class and say, "What kind of sound does this begin
with?" (thus making a point about "use" and "mention" and illustrating
that deictic reference need not cast a shadow), but he would not be advised
to lead the class out of the building, over a long winding path to the
roadway, point to a car parked at the curb, and ask what sound it begins
with. Illustrations—we presuppose—will keep their place no less than
people.

VIII

So far I have focused on the issue of cognitive presuppositions as reflected
in the syntax and prosody of utterances. But I hope it has become clear
that another phrasing would have been possible, one that directly ad-
dresses the quite fundamental issue which so far linguists have not been
able to resolve, though not for want of trying: namely, how, in a systematic
way, to get from what an individual says to what that individual means.

One attempt to solve the problem was initiated by John Austin as an
implication of his notion of "performative utterance," or "performative,"
namely, an utterance that does something. The tack was to consider the
condition which would render such performance null and void, that is,
"infelicitous." The assumption is that the effective execution of a perfor-
mative would have to satisfy a list of "felicity conditions."[21] Austin (1965,
pp. 14–15) lists six—a rather modest number to which to reduce the
subject matter of those who study face-to-face interaction. Later John
Searle (1969) refined Austin's effort to classify types of performatives (now
called speech acts) and made an effort to explicate the felicity conditions
(now partly covered by the term "sincerity conditions") presupposed by
some of them. This line of work has merged with an analysis initiated
by Paul Grice (1975), who recommends four maxims cooperative speakers
must be guided by if their utterances are to be maximally usable by
recipients.[22]

The short list of Austin-Searle conditions for the felicitous performance
of a speech act and the shorter list of Gricean maxims are presented as
culture and context free; they (along with the rest of "speech act theory")
have had a vast impact on sociolinguistics. As an analysis of speech in
context, the whole approach might strike the sociologist as somewhat

[21] Austin (1965, pp. 21–23) excludes from infelicities certain ways in which what appears
to be a performative really is not, in part because these "parasitic" forms could infect any
utterance, not merely a performative; but he admits a clear line cannot be drawn. I follow
Searle (1969, p. 61, n. 1) in including these etiolations.

[22] Be as informative as is required, but not more so; say only what you believe is true and
for which you feel you have adequate evidence; be relevant; avoid ambiguous or obscure
expressions or unordered presentation.

optimistic, if not silly. However, there is an important element in this scholastic foray into microanalysis which provides some useful suggestions as to the way the mind works in conversational analysis. For it is assumed that felicity conditions (broadly concerned) and conversational maxims will constantly be breached, causing the hearer to reread the utterance as an expression of unseriousness, sarcasm, understatement, rhetorical question, and the like. These "keyings" provide something like a systematic convention-based means for shifting from what is more or less literally said to what is meant, a presupposed interpretive repertoire that introduces much flexibility in the presuppositional bases of reference and inference. Now, although felicity conditions and conversational maxims may provide us with an inadequate means for determining when an individual is serious or cooperating as an interactant, the patent breaching of one or more of these rules may provide us with a culturally recognized signal that something unserious, skittish, or fey is intended.

In this paper I will be concerned with a further point, namely, what happens when these rereading rules, these presuppositions about presuppositions, are *themselves* breached. For when the "literal" (or, in Austin's terms, "locutionary") content of an utterance makes no sense in the context, *and* neither do conventional keyings, then a drastic interpretation must be made: namely, that the speaker is temporarily incompetent, or, if there is no corroborating evidence for that, that she or he is deranged.

Take, for example, anaphora. Its use is not only a right but also, betimes, an obligation. Speaker's failure to exploit available means for succinctness can lead hearer, and be meant to lead hearer, to look for an indirectly expressed intent, namely, that the speaker is being unserious, emphatic, sarcastic, ironic, distancing, overly polite, and the like. So, too, when hearer seems to fail to catch the speaker's obvious intent and responds as though the speaker's words were being taken literally; this also must be read as anger, joking, teasing, and so forth. However, when these "normal" ways of saying one thing and intending another, or acting as though speaker has been misunderstood, do not plausibly account for what is in the mind of the individual in question, then a second interpretive step must be taken, and another order of explanation must be sought. To wit: that the individual is temporarily incompetent owing, for example, to tiredness, preoccupation, inebriation, or narcosis. Or, if not this, then unhappily that he or she is strange, odd, peculiar, in a word, nutty. Either their mind is fully and fixedly not where it should be or they presume the other's is where it is not reasonable to expect it to be, or both. Each utterance presupposes, and contributes to the presuppositions of, a jointly inhabitable mental world, and even though such worlds last only as long as there is a warrant for a common focus of cognitive attention, one should not think one can go around failing to sustain them.

As with anaphora, so, too, with topicality. In many states of talk, an abrupt change in topic is to be read as a pointed thing, done to convey something about speaker's relation to prior topic that cannot easily be put into words without, for example, giving offense or making disclosures. Should such readings of breaches not be available in the circumstances, hearer can only surmise that speaker is temporarily not himself, or, failing that, that something is profoundly defective in his relation to what has been jointly entered into.

Perhaps a qualification should be added here. As suggested, in conversation persons are not obliged to adhere to the topic (even versions unanticipated by the initiator), providing they give a little prefatory excuse for failing to do so. So in some conversations, topic is more often evaded than sustained. But note that there are two senses in which mere excuses are not really mere. First, an excuse provides a warning that a topic change is about to occur and so warns the hearer that something outside the current focus is about to be placed in attention, and that a memory search may now be required in order to make sense of what the upcoming remarks will prove to deal with. Thus, "alerts" and "disjunct markers." Second, when speakers show they are concerned about norms with respect to topicality or relevance albeit not sustaining them on this occasion, their demonstration does indeed accomplish what appropriately relevant behavior would do, namely, give evidence that they are properly alive to the current situation.[23]

All of which leads me to hazard a definition of the felicity condition behind all other felicity conditions, namely, *Felicity's Condition:* to wit, any arrangement which leads us to judge an individual's verbal acts to be not a manifestation of strangeness. Behind Felicity's Condition is our sense of what it is to be sane. Well, of course, that has been argued for some years now. What is new (as opposed to given) is that syntactic and pragmatic analyses are to be seen as describing in empirical detail how we are obliged to display that we are sane during spoken interaction, whether through the management of our own words or the display of our understanding of the words of others.

IX

I have so far considered the bearing of some linguistic analysis on the question of what we take for granted when we talk to one another, and I have not been fastidious in pointing out the normative presuppositions

---

[23] In a useful paper, Baker (1975) rightly argues that prefatory apologies can be employed to mitigate any dereliction of conversational duty, including, e.g., the violation of any of the Gricean maxims.
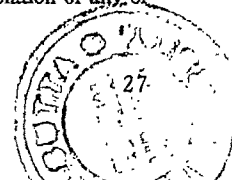
of some of this work. It is time now to reverse directions and consider what sociological analysis can contribute to linguistic formulations.

Given Marsha and John lodged in face-to-face talk, the argument has been that their cumulative discourse to that point, their jointly perceivable surroundings, and the knowledge each knows or assumes the other has brought to the encounter, can all provide understandings that are presupposed in the phrasing of a next utterance and without which the relevant meaning of the utterance might not be easily discoverable. These understandings can readily be seen as essentially *cognitive* in character, the only normative feature being the obligation to speak in an understandable way. However, if for a moment we think of these three sources of presupposition not so much as a resource to use in speaking laconically but as bases of constraint in formulating utterances, then it becomes reasonable to consider another factor: the moral norms of considerateness which bind individuals qua interactants. Delicacy, courtesy, modesty, politeness—these are the sort of attributes that are involved.

Take, for example, referential tact. Speaker is obliged not only to be alive to what is in his hearer's head—alive enough to find a phrasing that is succinct yet sufficiently detailed—but also to find a phrasing that is inoffensive and not indelicate in the circumstances. Meeting John on the street, his friend has the right (indeed, often the obligation) to ask him how his wife Marsha is and to use her first name as a fully specifying designation; John will know which Marsha his friend is referring to and will have a ready answer. Meeting John on the street just after John (it is known) has broken up with Marsha, his friend omits the nicety, and not because John would not know who was being spoken of. (However, now asking after Marsha can be in order if the phrasing carefully distinguishes the concern from the perfunctory kind, somewhat as it might were John and Marsha still a couple but she had just spent her first week in prison.) Our saying something that is considered to be salacious, intrusive, ill-timed, impertinent, or importuning is not to presume wrongly on what the other can easily bring to mind, but to presume in quite another sense, namely, to raise questions that the recipient might want to avoid considering with the likes of us. Thus, to say that there is a set of persons that John and Marsha *can* effectively specify to each other for the first time in a conversation by means of first name does not tell us whether they are free to do so. One would also have to know whether John and Marsha are in such a social relation to each other and to the mentioned party that this use of a familiar form of reference is thought to be appropriate between them. John, speaking to his and Marsha's seven-year-old, asks the child where mummy is, not where Marsha is, although the child will know perfectly well who is being referred to by

that proper name when, for example, she hears her father use it in talk to an adult family friend. This reference practice, then, is based not merely on what has been referred_to-in the talk, can be pointed to, or imported informationally; it is also based on the verbal license and constraint normatively binding the two participants. (Observe, although these constraints specifically condition speech, neither a syntactician nor a phonologist would have much to say about them.)

John can, of course, *refer* to these niceties using laconic expressions to so do, but then he would be taking for granted Marsha's knowing what he was talking about, and this is quite a different kind of assumption than that of his presupposing that Marsha will have (and exercise) these politeness norms in the uptake of his remarks. To presuppose another's understanding is not to presuppose another's capacity and disposition for considerateness in talk.

Now it should be seen that one's cognitive presuppositions about the interactional capacities of others present can become closely mingled with politeness understandings, so much so that on occasion the distinction may not be easily or even profitably made. For example, when John speaks to Marsha he cognitively presupposes that she can understand and speak the language he uses (much as he does that transmission conditions will allow her to do so effectively), but he also presupposes that she will be reasonable enough to put her language capacity at the disposal of the exchange he has initiated. And there will be similar presuppositions regarding the mechanics of encounters. For example, John will assume that when he follows conventions for summoning others into talk, they will allow themselves to be summoned. That once lodged in a state of talk they will give their attention over to it. That they will allow him to complete his turn without "unduly" interrupting him and will take a turn when the prior turn allows or requires it. That topicality will be respected. That a modicum of wariness will be maintained to deal with mis-framings should they seem to occur. That when the encounter is to terminate, all parties will have fair warning. And so forth. (Which is not to say—as Candace West reminds me—that when John is Marsha [i.e., when the person in question is female] these politeness presuppositions will not be qualified by still others regarding the relative weight that the two sexes can anticipate in cross-sex talk.)

More specifically, if the participants can assume that their purpose is solely to use talk in the rational, efficient, instrumental pursuance of some joint enterprise—and presumably this happens occasionally—then something like the Gricean conversational maxims (or rather, admonishments) will apply, establishing normative, not merely cognitive, standards for the exchange of relevant information. What we find in these cases is that

cognitive requirements for sustaining staccato, machine-readable communication are underwritten as part of the considerateness the participants owe one another in the circumstances.

Although constraints on the laconic formulation of utterances (whether due to what has not been provided in the text, in the local environment, or in joint experience beyond the situation) do not seem to be entirely of the same order as constraints due to politeness rules, there is one critical way in which they are similar: response to apparent violation. Most literally offensive utterances are not designed to be taken at face value— for example, as malicious or insulting or inconsiderate. Rather, as Lakoff (1973a) suggests, a framework of reinterpretation is applied, leading the recipient of the apparent offense to read it as jocularity, inadvertence, or another of our standard discountings. And just as with cognitive presuppositions, if the context supports neither a "literal" interpretation nor a conventional reframing, then some kind of deep mindlessness may have to be imputed.

X

I have argued that when you ask another what she thought of the movie last night, a variety of constraints can be involved. The question is not merely, How do you cast or frame your utterance so that it will presuppose all the information that is shared by all the participants but no more than is shared? (Although admittedly a principled account of that achievement would be useful to have.) Nor, as already suggested, is the question merely that of maneuvering among the never having been known, the recallable, and the already in mind. For these are all largely cognitive matters. The question is also that of social propriety—presuppositions and presumptions. And central here will be whether the putative speaker is acquainted with the putative recipient, and whether a state of talk already binds them or must be first established.

Consider this: You see a stranger at the movies whose appearance happens to stick in your mind. The next day you see this person on the street. You approach and without further ado ask, "What did you think of the movie last night?" No way. However, if the two of you end up next to each other in a queue that turns out not to be moving at all, and common disgust comes to effect a lowering of mutual barriers, then you can ask (in a properly tentative voice),

> Say, didn't I see you at the Regent last night?

The stranger may then reply, quizzically,

> Why yes, I was there.

You can then, for example, give an ironic grunt (as if to say that since the queue has made us act as though we were all here on purpose as friends passing the time of day, we might as well act out our forced appearances) and, in a quipping voice, ask,

What did you think of the movie?

the jocular tone providing serious evidence that you appreciate that a "straight" version of your question would be untoward.

But what if it is a friend you saw at the movie who apparently did not see you? Next day, on happening to meet, you ask (in greeting-adjacent slot and in a slightly taunting voice),

What did you think of the movie last night?

in which case you will be heard as having played a small trick, inducing in response something like a double take and a dutiful display of mock anger, as in

Hey! Were you there too? Why didn't you tell me?

this being an indirect allusion to the rule that what could have been part of joint experience should have been managed as such from the beginning, which, incidentally, would provide the basis for succinct references later. The unteasing way of going about it would have been to repair the joint biography (preferably in greeting-adjacent slot so as to avoid being trapped into appearing to have been ready to let the matter pass):

YOU: I didn't get a chance to say hello, but I saw you at the movie last night.
FRIEND: Really? Were you there? I wish I'd known you were going.
YOU: So do I. Incidentally, what did you think of it?

Or, alternatively:

YOU: I didn't get a chance to say hello, but I saw you at the movie last night.
FRIEND: Hey! That's funny. I saw you too, but there was no way to say hello.
YOU: Oh, no! By the way, what did you think of it?

Of course, if you saw your friend at the movie and your seeing was seen and acknowledged by a nod (and perhaps a pantomime display of wanting to come together but not being able to)—the "natural" way for things to happen—then, the next day during a chat, and at almost any interchange pause during the chat's course, you could jump right in with

What did you think of the movie?

For then each party would not only have had something of the same

31

experience to draw on, know that the other had, and know that this knowing was known to both (as Clark and Marshall [1978] suggest), but also know that this information state could be openly acknowledged—something, incidentally, that might not be felt were the theater a pornographic one. Nor, as suggested, should the situation of John and Marsha be taken to be entirely innocent here. Even if they went to the thing together, some issues would remain. For John to properly get off his "What did you think of the movie last night?" he and Marsha must be in circumstances where, for example, they are seen by each other to be involved in untaxing tasks that allow for the maintenance of a desultory side conversation. It is only under some such circumstances that a question of this sort would make sense. So, for example, upon both returning from the day's work and seeing each other for the first time in nine hours, John cannot march right into "What did you think of the movie last night?"—at least he cannot if his manner does not express that something like irony or hostility is meant. Nor if either has brought home guests can the topic safely be raised precipitously, for under lots of circumstances the guests would not be able to see how they could figure as relevantly present and could well feel they had a right to. At the very least, they would first have to be brought into the picture—both the interactional and the cinematic one.

In sum, then, one can ask, Given that you have something that you want to utter to a particular other, how do you go about getting into the circumstances that will allow you appropriately to do so? (The opposite question is of interest too, namely, How do you go about avoiding the circumstances in which you would be obliged to disclose something you would rather not?) Here, clearly, philosophy and linguistics must give way to sociology.

One answer seems to be found in a stepwise process. At this time, I can catch only a glimmer of a few of the steps:

1. If you are not acquainted with another and are not in a state of talk with that other, is there anything you can properly say flat out, without summons, warning, lead-in, or preliminary arrangement? Differently put, what are the circumstances in our society in which randomly selected individuals can rightly take it that they have the same thing in mind as the next person, and that this same thing allowably can be addressed without preamble?

   i. First, unacquainted's greetings. There are lots of environments in our society, even in urban areas, where unacquainted individuals may properly greet one another in passing. Having monitored the other's approach (and the other, yours), consciousness can be assumed to have been readied. Similarly, on arrival at a staging area—an elevator, an entrance

way, a waiting room—a newcomer may offer a greeting to a solitary stranger already there.

ii. You are standing next to someone whom you have never seen before. You are both watching what our view of life warrants our defining as a dramatic event: a three-alarm fire, an ambitious display of fireworks, a daring circus event, a very long line drive into the stands, an aerial display, precarious construction work. You have the right without preamble to half turn to the other and vent a response:

> Oh, God!
> Did you see that?
> Look at him go!
> Incredible!

to which the other properly responds by displaying that they have not been improperly addressed; they shake their head as evidence of similar wonderment, or possibly reply with a remark such as,

> I'll say!

Or

> Fantastic![24]

(These events occur during purposeful "official" watching; however, unforeseen circumstances can also provide warrant. Thus, should an airplane you are in very sharply jerk or lose altitude, you can unceremoniously exchange exclamations with an unacquainted seatmate, as you can with a fellow pedestrian when you both see that you have both witnessed a sudden near accident.)

Once the interchange is completed, neither of you may feel the need to continue to talk, although it is probably the case that the way has been made a little easier for a second exchange at some later point. You need not feel that any ice has been broken. But when either (or both) of you depart from your vantage point, a departure nod is acceptable. (Throughout social life, when departure is assuredly in progress, a show of bondedness can be afforded, presumably because there is then a guarantee that advantage cannot be taken of it.)

In brief, when jointly faced by what are to be taken as dramatic occurrences, any two individuals in the community can acceptably assume

---

[24] Individuals situated as herein described teeter on the edge of talk. A "free good" (see n. 29 for definition), e.g., can easily be requested or offered, this in turn giving either party something of a right to extend the crossover with small talk. ("Response cries" can similarly provide an opener, at least for the hearer.) Consequently, someone desirous of offering or requesting a free good might inhibit the impulse, not wanting to risk the possibility that the opening thereby given might be extended by the other into a sustained state of talk.

what is in each other's mind, can laconically bridge remarks to it, and will not be seen to presume in doing so. Incidentally, one might be hard put to discover what is the "new" in these expressions of wonderment. Note, too, that whereas the acts earlier considered presuppose a common language—or, rather, knowledge as to whether the language is likely to be common or not—the remarks here considered are almost independent of language. Given that "deeply human" events are felt to be involved, and that such events are felt to strike everyone similarly even though they may perceivedly be of somewhat different culture, and given also that no new information is in any case involved, it is understandable that language itself need not be seen as a barrier. Indeed, among Western people, there are nonvocal gestures which can function as fully equivalent alternatives in these settings. In any case, strangers to the local language well know that at least here they can communicate to the natives.

iii. A further possibility. You are walking along the street. You accidentally bump into someone. Both of you initiate a perfunctory excuse and continue on. (The guilty party, if there very clearly is one, has some obligation to be the more vociferous in verbal apology and to initiate physical remedial actions, if any seem required.) Again, no state of talk need be established in advance and no warning or preamble need be given for the remark. The reference is to something that can be taken to be in both parties' consciousness, and, being there, to be also a proper matter for address by either party. Interestingly, there is considerable variation across societies, and across classes within a society, as to what constitutes excuse-warranting mutual impingement. What is in our minds and can be properly taken to be in our minds during such moments may not qualify as worthy of entering consciousness in other groups. (For example, minor accidental touchings that elicit apologies in pedestrian traffic in our middle-class society routinely pass by unremarked in Brazilian streets.)

iv. Consider another scenario. You go up to a ticket window at the movies and, while fishing for your money, without further ado say to the ticket seller—who, as usual, is a woman,

> Two, please.

Following Chafe (1974, p. 112), one can say that your utterance presupposes that tickets are what you want, that the seller will have the possibility that you want tickets in focal consciousness and therefore will not have to canvass her memory to figure out what in the world you have in mind, and that certainly she is obliged to accept the request as appropriate. All the ticket seller will be waiting for, presumably, is a strategic bit of information not otherwise available in the situation, namely, how many. (As a matter of fact, microecology will allow the seller, if she is inclined, to predict how many tickets you want, so your informing words

can be merely confirmatory.) Further, the ticket seller will appreciate that she is not being asked to *give* you tickets but to *exchange* them for money.[25] What is special about this scenario, however (and unlike Chafe's [1974, p. 114]), is that no prior state of talk is presupposed; just by walking up to the window, that is, by lodging yourself in current-turn position—the "service post," to use Marilyn Merritt's term (1976, p. 321, n. 14)—you have a right to start flat out with words regarding tickets.

Part of the job of ticket selling, then, is that the performer is obliged to have in mind what it is that will make sense out of cryptic utterances regarding tickets and so orient to any unacquainted other who might approach the service post. And this obligatory matter for consciousness does not come from prior utterances in a conversation but from an institutionalized service arrangement and the probable transactions that will be engineered in its terms. Indeed, the whole transaction can be, and frequently is, consummated silently. Observe that, in some communities, if you treat the transaction as one that requires you first to initiate a state of talk before saying how many (as with, "Excuse me, Miss, could I have . . ."), then the seller may well look to her own behavior and your tone of voice for evidence that a sarcastic reminder of her duties is intended, and (we will soon consider) if she fails to find any such evidence, she may think your head needs tightening.

v. A final matter. The general rule in our society against opening talk with "strangers" seems to be understood as one that should not be sustained too pointedly. The injunction against talking to strangers is merely one part of what we presuppose, another part being that acts which glaringly portray distaste for contact with another should be avoided. Thus, if fortune brings two adults into close and exclusive co-presence for a considerable time, as when an elevator jams with only two riders on it and they are unacquainted, then it may be felt to be improper to refrain from exchanging a few words, in this particular case especially (but not merely) because each can assume the malfunction will be on the other's mind.

2. I have so far touched on some of the social circumstances in our society when an individual without first initiating a state of talk can address an unacquainted other on the basis of what can properly be assumed to be in the other's consciousness, and being there, to be properly available for address. This speaks, of course, to what it is that certain individuals may be *obliged* to accept from unacquainted others by way of unceremonious address. A marginal matter should be considered next, an institution that plays a special role in communication life and has a

---

[25] We describe the utterance "Two please" as a *request* for tickets, but in fact the only thing being requested is that an exchange of tickets for money be initiated now. We mean, "Could I buy two tickets now," but we do not say it.

distinctive role in regard to what it is we must be prepared to bring to mind. I refer to the "summons," that is, a message conveyed to someone otherwise engaged, directing him to address himself to the speaker in the matter of their establishing a state of talk between them.[26] One deals here with a relationship that pretty much extends to all members of a society, linking every possible pair, and each member both as giver and sender of the signal. For, at least in face-to-face life, it is always possible to imagine *some* circumstances which would provide warrant for communication overture between *any* two individuals within hailing distance of each other.[27] Thus the signals:

> Excuse me.
> Pardon me.
> Sir/Madam [with rising intonation].

Of course, we are not obliged to maintain the possibility of overture ready in our consciousness, merely to bring it to mind when we hear an appropriate signal, especially when followed by an appropriate orienting action on the part of the signaler. Observe, once again, that the use of such signals implies that the user appreciates that we are not to be brought into talk unless given a moment's notice and, incidentally, that we have the right to defer the opening by employing words that request the summoner to place himself on hold while we complete an urgent task. (There is always the possibility of not responding to the summoner at all [or not answering the phone], in which case consciousness is likely to be full of the overture, although no confirming action is forthcoming regarding it.)

3. Summonses aside, consider some of the circumstances in which individuals may feel that to make a statement they must first establish a state of talk to house it in; and, once that is done (or if it already has been done), that an "alert" should be employed, namely, some sort of prefatory warning that information outside of immediate consciousness may well have to be brought to mind.[28]

[26] "Summonses" were first considered systematically (and under that label) in Schegloff (1968). Sacks (1972*b*, pp. 344–45) used the term "ticket" to refer to warrants for unceremoniously initiating a state of talk. Nofsinger (1975) examines summoning into a state of talk and applying for the right to tell a little story in a state of talk already established, considering both under the term "demand ticket."

[27] There is a summons cue that also functions to gain the attention of the head speaker in a meeting, the consequences of which will be a speaker/addressed-recipient exchange within a larger whole. However, such turn requests seem to carry little of the force of the primary kind. After all, those seeking to be ratified as next speaker by head speaker are already engaged in a state of talk—if only as unaddressed recipients—and the person they are trying to make contact with is not in transit beyond their ken.

[28] Alerts can be compared to what Jefferson (1978, p. 221) calls "disjunct markers" and "embedded repetition," these being two sequenced devices for gearing a storytelling into a conversation. Disjunct markers, of course ("Incidentally, . . . ," "By the way, . . ."), are also

i. In public places, passing strangers have a right to breach the rule of nonsolicitation to request certain "free goods" of each other, among which is information about direction and places.[29] Of course, there is no obligation on the part of the citizenry to possess the information, merely to countenance a request to search their memory for it. (The obligation of the asker is to approach the potential informant in the situation who might be considered least threatened by the overture.) The request is often prefaced by an excuse for initiating talk and an alert regarding the search—a two-step arrangement:

> Excuse me, would you happen to know. . . .

Through the same route, passersby can be asked for the time and, under certain circumstances, change for a parking meter or telephone, in which case the free good that is solicited is something other than information, yet something that falls within the domain of matters every citizen must be prepared to address. (Any stranger could be asked what day it is, but few are because it is also assumed that every normally oriented citizen should know that; only very young children can properly tap this resource without doing sheepishness. The *date,* however, can be asked [half apologetically], e.g., whenever the asker can be seen to be filling out legal documents.)

Balancing off the right to elicit a number of free goods, there is the obligation (which can also be a right) to volunteer some others, for example, informing someone that they have dropped something, offering

---

employed to show recognition of, and apologize for, other threats to coherence. A general treatment of "speech markers" is available in Schiffrin (1982).

[29] I use the term "free good" to refer to those things perceived as ones a possessor can give away at relatively little cost or inconvenience even though they may be urgently needed by the recipient. (I do not, therefore, mean sunshine or public tennis courts, although these goods could also be called free ones.) Free goods, as here defined, provide an interesting source of social solidarity. Their range varies greatly according to ecology. For example, although fellow pedestrians and fellow motorists ordinarily have quite limited resources of this kind, those who pass each other on isolated desert roads in effect have greater resources and are (in the Southwest) more likely to initiate the offer of them. (Similarly, close neighbors in otherwise isolated areas are well situated to supply each other with a wide range of free goods, and, correspondingly, have a good reason to sustain friendly relations.) All of which is not to deny that free goods are often withheld from those we define as beyond the pale—persons who by their appearance and manner suggest that their need is chronic or even simulated, and that they are systematically exploiting pedestrian charitableness for what can be accumulated through repeated requests. Thus, an affluent-looking stranger may request or be offered an appreciable free good on the assumption, presumably, that no "working" of the system is involved (and, of course, that such a person would be in a position to offer the same level of help to another on a different occasion). A full consideration of the implications of free goods in dealings between cross-sexed strangers is reported in Gardner (1983). A consideration of direction giving is provided in Psathas and Kogloff (1976).

to help with a cumbersome task, and so forth. Here again, a state of talk may well be initiated first, and a warning that the recipient of the summons into talk should check to see what might be a likely reason for the overture. (Apparently you are always supposed to have your mind sufficiently in shape so that if you are told you have dropped something, you can quickly bring to mind that you had this something to drop.)

ii. Pedestrian obligations with respect to being called into talk and being required to bring to mind certain pieces of information, engage in bits of action and the like, are bilateral; individuals who can draw on this corpus of knowledge and action must also be able betimes to make theirs available. There is a contrast here with persons who acquire a specialized role as information givers (and performers of free services), whether formally, as in the case of policemen and information clerks in shops, or informally, as when a news vendor or filling station operator comes to be routinely asked for directions, or a drugstore that is near a bus stop becomes a source of change.

iii. The notion of free goods leads immediately to the issue of what goods, and how much, are to be considered free, and this whether on the part of passersby or specialized providers. A more interesting issue, perhaps, arises when the good is information. Here knowledge, not cost, is felt to be involved.[30] Informational preserves aside, it is as though before making our inquiry, we have to judge how "reasonable" it would be for the addressee to have the information we want, and in making this judgment to conclude that our request has a "respectable" chance of being satisfied. For to ask someone for information they could not "reasonably" be assumed to have is a perceivedly disoriented (and disorienting) thing to do. Which is also to say, of course, that by nicely expressing that we realize how unreasonable our request is, we can get away with making it. Thus locutions such as,

> You wouldn't happen to know, would you, where [what, when]. . . .

Contrariwise, the more reasonable the request, the less ritual labor required to ask it. But, of course, these arguments bring the question back to the issue of what is defined as "reasonable." It seems that several principles are involved, all speaking to the issue of the recallable or bringable to mind, of standard stocks of knowledge, not to what ought to be in consciousness at the time.

Take, for example, location information at various removes from the

[30] In politics, business, and courtship, information can, of course, be strategic and as such anything but free; but this sort of information is not likely to be asked for as if it were a free good. More generally, information about self is a form of personal territoriality and an important ceremonial adjunct to the organization of social relationships (Goffman 1971, pp. 38–40).

point of inquiry. There is the assumption that persons whose work or domicile fixes them to a given point will have detailed place knowledge of the immediate vicinity. What is asked of suburbanites, news vendors, shopkeepers qualifies here. Of course, "immediate vicinity" (like "reasonableness") itself requires explication, drawing on presuppositions instead of describing them. There is, for example, the question of grid: a salesperson in even the largest department store can be reasonably asked where any particular department is, a filling station attendant where the access point to any neighboring highway is; the first case is on a finer grid than street-place information, the second on a coarser.

In contrast to those perceived as fixed, we have police and cab drivers. They are obliged to take as reasonable a request for information about any "place" in their city; moreover, should they not "know" where the place is (in the sense of being able to direct you or take you), they are under some obligation to recover the facts from a book or a phone call. The citizenry itself can be asked for directions to the city's "leading" grid of downtown streets and landmarks, but hardly for more. Apparently another principle is this: If you feel that your specific request is questionable, you can select a larger segment which (you happen to know) includes your point of interest and ask about this broader area, the assumption being that the greater the distance from your informant's familiar location, the more coarse-grained his grid information. (I might add that if you ask for a place that is within arm's reach, your informant will have to exercise a special tact in informing you; he or she may give, for example, a little sympathetic laugh to show that this sort of blindness is not being taken for what it might be.) So all in all, if you want to fine tune your requests concerning location, it would be a good idea to know in advance just what the answer is going to be.

Interestingly, in replying to a request for location information, the respondent may have to resort to use of other location terms which mark out the route to the asker's destination, but, of course, an asker who does not know one point may not know others on the way. The respondent may have to engage in a probing search procedure, employing "try-markers" to test out various landmarks until one is found that the asker is familiar with. (It need hardly be added that with or without such side sequences the informant's efforts may be insufficient; often she will not be aware of, and hence cannot advise on, all critical decision points that she herself negotiates when following a particular path.)

iv. Return now for a moment to our movie house ticket seller. One finds, of course, that she is not merely exposed to persons who want to buy tickets. A range of other possibilities exists. Potential patrons can come up to the wicket to elicit information about the film: how long will it run, what time does it start or stop, does it have subtitles, is there a

cartoon, and so forth. Although these requests for information do not tap what the ticket seller can be presumed to have in the center of her awareness, the presumption can be that she is certain to have access to such information, is obliged to divulge such information, and that a deep search will not be required. Hardly anything by way of an alert need be offered her.

However, with some other matters she can be approached on, ceremony figures differently. As suggested, it can be expected that she knows where particular shops and buildings are in the immediate vicinity of her workplace, and that it is permissible to query her on such matters. Similarly, one can permissibly ask for change for a nearby phone or parking meter. In all these cases of solicited free goods, there is no obligation on her part to provide the service, merely not to take umbrage at its request. Indeed, in the matter of the time, anyone can ask and everyone can expect to obtain an answer.[31] But in all these cases, the asking is often prefaced by a polite alert; certainly such a courtesy is understandable as fitting when it does occur.

I am suggesting that requests that call for dredging or bringing to mind are more likely to be embedded in a perfunctorily established state of talk (along with a polite initiation and close-out) than are requests for tickets. The distinction between what is in consciousness and what is merely recallable is thus reinforced by ceremony; if the ticket seller need only have a matter available to recall, then a request concerning it may well be felt to require first that a state of talk be established. Thus, there is a protective structure to her exposure.

Clearly, the little scenarios we have so far dealt with involve individuals who are allowed to address one another on a very narrow range of topics, all other topics ostensibly forbidden. You are not entitled to ask a ticket seller flat out, without preamble, whether her hair is natural or not, what she thinks about her mother, and a host of other matters she can call to mind and routinely will when talking to intimates. And the same sort of reasoning that applies to asking for information applies to the practice of initiating the divulgence of it, namely, to "reportables." You can tell the ticket seller that you are dying to see the show, and while this might be judged as a little oversociable, it is perhaps forgivably so. But if you

[31] Not quite everyone, of course. For example (as Bambi Schieffelin has suggested to me), although it is deemed reasonable to wait in line in order to buy a ticket, to do so to elicit a free good might seem odd, perhaps because a free good, after all, is the sort of thing that almost anyone could supply. If the askers are to show they are alive to what they are doing, any marked inconvenience to asking one source should cause them to seek another. Indeed, even in matters cinematic, an asker will be constrained to display a "sense of proportion": to wait in a long line merely to find out who is starring in a film could strike witnesses as strange.

chose the moment of contact with a ticket seller to tell her directly about having to take the car in tomorrow to get a new muffler, that would be considered grounds for imputing strangeness. Obviously, what is reportable "on contact" to a spouse may not be thus reportable to a stranger. And yet, of course, it appears that in certain circumstances strangers can "strike up a conversation"[32] and end by exchanging extensive intimacies, not merely "small talk."[33] And if you are at a ticket wicket that is not busy, and there is a displayable reason for hanging around, you can acceptably lay the groundwork that will supply the perceived relevance and conversational license that is required to allow either her or you to properly report, or inquire about, matters ordinarily precluded.

4. So far I have considered only the circumstances in which we ourselves seek to obtain or reveal information, the question of who in particular we are led to address to this end being of second importance. Omitted, then, have been all those occasions when the asymmetry is on the other foot, and we find ourselves *required* to divulge information about ourselves.

Servers of all kinds have the right to ask those they serve for pertinent biographical information, psychotherapists presumably marking the extreme in this regard. (To seek a service, then, is to expose oneself to questioning.) Agents of social control—police and teachers, for example— exert a somewhat similar claim, except that here the informant may not have sought out any assistance and correspondingly may *more* resent the right of the askers to ask what they do. Rights here, of course, are unilateral. In any case, in coming to a server we ready our thoughts so that the dredging process will be facilitated. The presence of a social control agent may bring certain topics to mind also—commonly ones we would rather avoid.

5. When one shifts from contacts between the unacquainted to those between the acquainted, especially the "socially" or not "merely" acquainted, then, of course, matters get considerably more complicated. Obviously, each party to the relationship will be obliged to recall on sight

[32] For a conversation to have been "struck up," we apparently assume that the participants had been unacquainted (so that to speak of "strangers" doing it is redundant) and that circumstances had positioned them close to each other with evidently "nothing" but waiting to do.

[33] Small talk is itself of considerable interest. Topics such as the weather, inflation, television shows, unsafety on the streets, and the like can be seen as "innocuous," meaning here that no offense is likely to be taken by anyone addressed on the matter. But surely more is involved. It is as if society had set aside certain topics as ones that everyone is expected to have an opinion on, that are to be held in relative readiness and to be brought to mind quickly, that are not defined as part of anyone's protected lore (and therefore can be addressed by social inferiors), and that are seen as requiring no special competence or experience. Small-talk topics, then, mark a socially sanctioned opening between minds.

the name of the other, along with some biographical details, for these
may have to be displayed should a greeting occur. Recognizing that certain
past experiences, and consequently certain familiarities, will be shared,
a base exists to be drawn on anaphorically and deictically. And as sug-
gested, presumably the more fully involved the two individuals have been
with each other over time, the richer and more current will be the joint
account of known-to-be-known-about available for allusion in setting up
an utterance. For example, if Marsha and Mary are adult close friends,
then each is likely to know, and be expected to know, critical matters
about the other's close others, enough so that each can efficiently report
stories involving her own others. I might add that ceremonies of greeting
and departure in our society often require reference to these "other per-
son's others," this being a case where what *can* be asked about *ought* to
be asked about. Such a rule underlies the insult offered (in *Pride and
Prejudice*) by Lady Catherine when, sorely vexed with the marital am-
bitions of Elizabeth, she departs into her carriage saying:

> I take no leave of you, Miss Bennet. I send no compliments to your mother.
> You deserve no such attention. I am seriously displeased.

And between close, long-standing workmates or housemates, each party's
preacquaintance past may be reviewed and assimilated somewhat to joint
biography, again so that (or, at least, with the effect that) a recallable
framework will be available for anything either party might want to say
involving reference to the past.[14] Correspondingly, there is the important
obligation (and right) to update our associates about any change in our
life circumstances—the "closer" the relationship, the more quickly the
information is to be imparted and the less dramatic this information need
be. In consequence, one's intimate others will always be properly oriented
to receive reports and make inquiries.

In brief, what we think of as a relationship is, in one sense, merely a
provision for the use of cryptic expressions, a provision of what is required
in order to allude to things economically. Certainly our obligation to keep
the names of our friends in mind, along with other pertinent social facts
concerning them, is more than a means of celebrating and renewing our
social relationship to them; it also ensures a shared orientation for ref-
erence and hence talk whenever we come into contact with them. What
affirms relationships also organizes talk.

6. With this biographical perspective in mind, turn finally from mere
and close acquaintances to the closely intertwined: back to John and
Marsha. They have come into the kitchen for lunch. For a while they

[14] The central treatment here is Berger and Kellner (1964).

will maintain an open state of talk, that is, the right but not the obligation to utter words. In consequence, there will be longish silences interspersed with short spurts of talk. And not a few utterances will receive no other-person uptake of any kind:

> JOHN [looking in refrigerator]: Marsh've ya seen the mustard. . . . Oh! here it is.
> MARSHA: . . . .

The general question, then, is, What can be reported on, and what can be inquired about, with what kind of prefatory work, if any, and what surface structure evidence will there be of what sorts of matters that are taken for granted?

First, there will be the domain of matters that can be addressed flat out without summons, alerting, or ground laying—in effect, matters taken to be already in consciousness and in addition permissibly addressable. Included here is "course of other's action." Marsha can see what John is about (getting more coffee, rinsing a glass, going to the refrigerator), and in the setting is licensed to take such doings as the frame of reference for flat-out remarks. (To help in her tracking of John, Marsha may use John's progress grunts and ostensibly self-directed comments which he may employ in executing his actions.) Observe, John does not have to see that his course of action is seen; he can have his back turned to Marsha, but provided the kitchen is small enough, his doings will provide a graspable tracing of what Marsha has social warrant for assuming is in his mind. Thus, as he gets up to get more coffee, she can laconically say, "Me, too." Further, if the pair are sitting or standing facing each other, Marsha can assume that John is somewhat tracking her course of action (and vice versa), and flat-out requests for specific help or offers of specific help are possible. Indeed, a considerable amount of giving help and getting it will go on without help from words of any kind—enlightened collaboration in spite of total ellipsis. Let me emphasize that one deals here, of course, with social definitions. Thus, in surgery the scrub nurse is obliged to stay "in" the course of action of the surgeon, so that without raising his eyes he can hold out his hand and get what he wants by nakedly naming it; however, she cannot expect him to similarly track her doings. (My pronouns presuppose that the traditional sexual division of labor still typically obtains in surgery; this parenthetical comment, however, involves a statement in this regard, not a presupposition.) Children, to take another example, often feel crowded by their parents, feeling that they cannot complete an act themselves without helpful intervention, both verbal and physical. On the other hand, strangers on the street need to be in fairly dire straits—but not too dire—to allow for unasked-for intercession. (I might add, if you stand at an intersection and look very, very ponderingly

in both directions, you may be able to get someone to volunteer the question,

> Are you looking for someplace?

But only a very pointed enactment on your part is likely to give Good Samaritans a feeling that they can chance asking before they are asked.) A man seeing a woman (even one not known to him) reaching for a match to light a cigarette can indeed proffer a light; but the reverse still carries more implications than most women might want to risk in the performance of a courtesy. A quarter of a century ago on the Shetland Islands, any man doing outside work would be likely to be asked if he wanted a hand by any male (acquainted or unacquainted) of crofter status who happened to find himself standing about in the vicinity. We do not have this expectation, although when cars get stuck in the snow, volunteers may leave the sidewalk to lend a hand, especially if there has been an extraordinary storm and citizens for a time feel the good-humored will that crises can generate. In other words, appreciating another's course of action is not enough. Helpful intervention on the basis of this cognitive outreaching may require obligation or at least license.

Now return to John and Marsha and their open state of talk in the kitchen. In addition to course-of-current-action as a legitimate window into the consciousness of the other, there will be a set of events entirely external to the current social situation which John and Marsha can assume are salient enough to mutually address without a topic alert or even a cue that talk is to occur. Marsha can break a long pause by saying,

> They seemed to have a good time.

She knows, and correctly, that John will understand she is referring to last night's guests. (Try saying that to a ticket seller.) But, of course, the notion of saliency is only a gloss; one needs a principled analysis. For example, if John and Marsha know they both know that they expect to be involved in an undertaking later that day, it need not be very much on their minds for one to say nakedly into an open state of talk without any lead-in,

> What's-his-name isn't coming 'til two.

In this case a name substitute has been employed that stipulates gender and nothing much else, much as a personal pronoun might, except that here there is recognition that initial mention ought to be a name, not a pronominal substitute.

Further, if a proposed topic (or rather a "mentionable") is perceivedly speaker centered, in the sense of being recognizably relevant primarily to speaker's own involvements, then a special set of alerts seems to be

employed in and out of established talk, as if to warn hearer that he or she is going to have to mobilize memory and orientation regarding a yet unnamed theme:

> *Geese* . . . I waited a long time at the dentist's.
> *Wow,* I didn't sleep much. I feel lousy.
> *Ya know,* I shouldn't have bought that sweater.[35]

John and Marsha are also likely to employ utterance-initial prosody and utterance-initial vocatives in a manner that informs hearer immediately whether he or she is being asked to become a sustained sympathetic listener or merely a one-fact informant—someone required to quickly complete a single interchange parenthetical to two ongoing unrelated courses of action.

Also, there will be the range of matters that John and Marsha can introduce into talk by employing standard locater formats, for example:

> Ya remember that green shopping bag we had?

May I add that what one party to the relationship has a right routinely to report to another (given the help of various alerts and the proper activity setting) is not an incidental aspect of what a relationship is, and a far trickier matter than what one party is obliged to report lest the other think that secrets are being kept or that third parties are improperly finding out first. The close study of such reportings takes us right into the banal interiors of intimacy (that domain of which psychotherapists are the absentee landlords), which might account for the studied neglect of the topic by students of talk.[36]

It might be tempting, of course, to think that there are some relationships, such as the marital one, that allow each member to report anything to the other and inquire into anything about the other. But, of course, there are not. Certainly with intimates each is likely to know enough about the other to know what the other is sensitive to (as we say) and cannot be addressed on except in a roundabout way—for example, only after relevance has been generated by apparently fortuitous events, including a buildup of what can be seen as unconnected mentionings. If the intimates are cross sexed, then there are certainly matters, especially

---

[35] I have also heard "Ya know . . ." used as an alert in small talk among the barely acquainted, the apparent function being to ready hearers for an obiter dictum whose level of generality departs sharply from the conversational particulars that surround it.

[36] Labov (1972, p. 370) is almost an exception: "To identify the evaluative portion of a narrative, it is necessary to know why this narrative—or any narrative—is felt to be tellable; in other words, why the events of the narrative are *reportable.*" Joel Sherzer reminds me that reportability is very much a cultural matter, citing Cuna Indian discourse, where, in the replaying of experience, incidental details are routinely reported to a degree we would judge excessive.

ones defined as quite minor, that each can more easily report to an acquaintance of own sex than to each other. And even when the intimates are of the same sex, and the same everything else that might make for license regarding reportability and inquiry, there will be matters defined as so trivial and fleeting as not to qualify for mention. Again, to closely describe these limits is to closely study the social relations of the people who draw them.

A final comment. Given the speaking arrangements described for John and Marsha, it seems reasonable to raise the question of "misunderstandings." We tend to assume that, although persons who do not share the same communication code are prone to genuine misunderstanding of each other ("genuine" in contrast to the many conversational moves based on simulated misunderstanding), intimates will at least be spared this source of divisiveness. But if anything, the opposite is true. After all, the domestically entwined can draw cryptically on own and other's course of action, and on a mutually appreciated agenda of impending involvements (both individual and joint); further, in our world they are likely to have ritual license both to initiate a spurt of talk with minimal forewarning and to intrude disjunctive topics with little ceremony during already established states of talk. It is to be expected, then, that in this relationship context, recipients will on occasion be in doubt about the reference of utterances and even mis-refer what they hear to a context not intended. And indeed ambiguity and misunderstanding, albeit short-lived, abound in the kitchen, on long drives, and other seats of domestic laconicity. Precisely here, where Marsha feels she can relax and communicate elliptically about "anything" that comes to mind, John may have to invest a certain amount of work in order to keep up with the changing contexts of reference; and so also Marsha with respect to John's utterances. I might add that the properties of these brief domestic hitches in communication have not yet been considered systematically.[37]

* * * * *

In sum, how an individual in talk with us, acquainted or unacquainted, intimate or stranger, can properly lead up to an otherwise intrusive query or a revealing report has never been closely studied; nonetheless, what is involved leads one back to a consideration of the workings of the new and the given. In shifting from considering information about directions or theater tickets to information about oneself—one's social attributes and one's prior experiences—the awful dimensions of the perspective emerge,

[37] A beginning is provided by Garnes and Bond (1975) and Grimshaw (1980).

namely, that conversationally speaking, we are all information storage drums, and for every possible interrogator, there will be an access sequence that allows for entrée to the files. The unilateral access that the police, government officials, credit bureaus, and professional servers of various kinds have to selected stretches of our biography has been considerably pondered, if not studied, by social scientists interested in the question of privacy—although here not particularly in terms of what this sanctioned access implies about the social organization of what people are ready to bring to mind. But I do not think much is known about the distribution of circumstances in which we volunteer reports on ourselves to those whose interest in what we report, incidentally, may extend only to the question of when we are going to finish our turn, so that they can begin their own reporting.

I want to note finally that what has been considered so far is, as it were, a hearer's-ear view of talk, namely, what it is hearer will be able and ready to respond to. Indeed, with small talk, speakers also take this view, rather fully determining what they say according to their view of what it is the hearer can be alive to. But this view gives too much weight to local determination. Speakers have the task of making what they say decipherable and relevant in the circumstances, but why they should have hit on one such particularity instead of some other one equally tailored for the recipient need have nothing at all to do with the latter. For example, John's glance may happen to have fallen on something that reminded him of some preencounter concern he now elects to mention; or something someone else says may fortuitously provide this associative link. And this connection may be totally inaccessible to Marsha. Thus, following an alert by John that breaks a period of silence, Marsha often will not be in a position to know what in the world John is going to say, although, of course, she can be fairly sure that after he has said it, the matter will turn out to qualify as a mentionable between them. (That is why the term "alert" was introduced in the first place.) So, too, in a conversation that is well under way, Marsha's utterance of a topic shift excuse ("Not to change the subject, but . . .") can warn that something not predicted is to be aired. The precise timing of such a shift can be locally determined, and the phrasal warning sufficiently stereotyped to call a "disjunct marker," but even so, just *what* the excuse maker then selects to report on can be a somewhat open matter from the perspective of the listeners. Indeed, speakers can make a point of attempting to conceal the reason for an utterance (as when they "fish" for a compliment, or employ a phrasing that disguisedly presents their real concern as one item in an innocuous-appearing series of questions), and although such an effort takes for granted that utterances will be assessed for the apparent

. reason for making them, it also assumes that in the last analysis assessors will be cut off from the information required for consistently faultless judgment.

XI

I close with a call to arms. To utter something and to not disconfirm that we are sane requires that our saying be heard to draw appropriately on one array of presuppositions—that sustained by our hearers—and avoid being heard to make others—those which are not, although they may be by persons not present. Responding to another's words, we must find a phrasing that answers not merely to the other's words but to the other's mind—so the other can draw both from the local scene and from the distal, wider worlds of her or his experience. Sociologists recently have not been very helpful here: they reiterate the proposition and provide illustrations. But there is no *analysis* of the taken-for-granted. No framework. Not even a simpleminded classification. The consequence of presuppositions for the surface form and prosody of utterances, on the other hand, has been considerably illuminated by linguists, along with the text-interpretive practices and the repertoire of keyings. The bearing of acquaintanceship and close ties, of the generation and intentional construction of joint biography, of being or not being in a state of talk, of the various locaters we employ to provide a framework for the statements we want to be in a position to utter succinctly—all these critical matters have been little studied. Behind all this, and linking these themes together, is the socially prescribed place of what is taken to be the operation of the mind. A question of who can say what to whom, in what circumstances, with what preamble, in what surface form, and, given available readings, will not be thought mindless in doing so. A question of what we can say and still satisfy Felicity's Condition.

No doubt the most closely examined case is that of "conversation." Anaphora and topical continuity are two examples of what has been considered in the literature. More generally, there is the matter of dialogic or sequence constraints. Given a turn at talk, a next turn will be required and will be examined for how it displays relevance to the now prior one. All kinds of discontinuities will be possible because all kinds of excuses will be available that establish respect for relevance even while failing to provide it substantively. Even one's failure to excuse a failure of uptake will be caught in the interpretive net, being read, for example, as a pointed comment on what has not otherwise been responded to.

But if Felicity's Condition is now best documented within conversations, this should not lead us to worship particularistic structures. To be sure, when John directs an assertion or question to Marsha, and Marsha

responds by remaining silent, or changing the topic, or turning from John to direct her own opener to Mary, Marsha's act can be perceived by all three as a behavioral comment, a reply in effect. But analytically speaking, to say that in context no answer is an answer is simplistic. Information derived from Marsha's failure to address John's utterance verbally, that is, canonically, is information given off, not given; it is (on the face of it) expression, not language. As such, this move figures in a special way in regard to what Marsha can be held responsible for having communicated. Whatever John says she meant by her apparent failure of uptake, Marsha has reserved the right to claim otherwise. For she has not *said* anything; she has merely placed participants in a position to draw varieties of conclusion. Should Marsha provide an explicit reply to John's utterance, she can still decline to accept particular readings of its hinted, tacit elements, for these, too, will be expression, not language.

It is true that prior turn is very likely to provide some of the context in terms of which current utterance will be interpreted, of which condition current speakers must demonstrate that they are cognizant. But on the same grounds, prior turn can never be the only such condition current speaker will be required (and allowed) to employ as a frame of reference. In speech act terms, respondents must address themselves to the illocutionary force of what is said to them, and to appreciate this force they will have to have access to a vast array of biographical and cultural understandings—these alone allowing them to make sense out of cryptic allusions. Moreover, matters not even alluded to in prior turn may be available for address by respondents, for in the last analysis it is the situation and circumstances of the prior speaker as these interact with the situation and circumstances of the respondents that the latter must address, the former's speaking merely providing the occasion for doing so.

Further, a conversation is only a conversation after the fact, this being the point when a student can be assured that a continuous stretch of examples can be culled of current turns being followed by contingent next ones. Occasions of one person's talk that lead to no other person's talk perforce fail to end up in the collection. The frequent necessity for speakers to continue beyond a first "transition relevance place" and take in their own washing for want of any volunteer is perhaps not given enough weight. Stops and starts and fitfulness generally are underplayed, as are long stretches of easy silence punctuated by unanswered remarks.[38] And the dynamics of participation are neglected, namely, the formation of subordinated communication (whether in the form of collusion, innuendo, or open byplays), the occurrence of dual or pivotal participation, the

[38] It should be noted that this bias is not a necessary result of the Sacks-Schegloff-Jefferson turn-taking model (1974), for the initial formulation leaves ample room for lapses in talk.

49

movement of participants from one adjacent conversation to another, and the subdividing and recoalescence of encounters (Goffman 1981, pp. 133 ff.). Most important of all, the sense in which current utterance is conditioned by immediately prior turn's talk—when, indeed, there is such talk—does not speak to the many elements of the same current utterance that are not determined in any way by prior turn (or prior utterance in the same turn), yet are nonetheless determined in ways that satisfy Felicity's Condition. In any case, an account of second utterances in terms of their contingency on a first leaves unexplained how there could be any firsts; after all, from where could they draw their design? Conversation could never begin. Or, once begun, would be one utterance away from the end. Tails would know how to wag, but there would be no dogs.

It is true that utterances must display that their makers are appropriately alive to the circumstances *in* which they make the utterance (as well as, of course, to the world *about* which they make an utterance), and must do so at their makers' peril, demonstrating that the minds of the latter are informed by exactly the required presuppositions. And true, too, that speakers can use these constraints as a resource (indeed, are obliged to), allowing them to employ efficient references with the understanding that what they must be mindful of is exactly what their listeners will be mindful of, too. But this informs us about the circumstances in which words are spoken, and these turn out to be very much broader than the circumstances in which conversations are maintained. This is neatly evident, for example, when two individuals are jointly engaged in a physical task and one is directing the other. "Now loosen the other one the same amount" is a deictic utterance whose required context is an immediately prior movement of a wrench, not a larynx, the resulting sequence melding turns at loosening and turns at talk.[39] There occurs an interchange of acts which are sequentially contingent and satisfy Felicity's Condition, but in a fundamental way the interchange is not conversational. A mutually ratified joint focus of attention is here sustained on and through a physical task, a contributory resource for which is an open state of talk.

Nor (it should now be plain) was it right to define Felicity's Condition restrictively in terms of verbal acts. Speech need not figure even in a reduced way for Felicity's Condition to apply: the general constraint that an utterance must satisfy, namely, that it connect acceptably with what recipient has in, or can bring to, mind, applies in a manner to nonlinguistic acts in wordless contexts.[40] These acts, too, insofar as they can be perceived by individuals in the vicinity, will have to be styled so as

---

[39] A useful analysis of reference and task activity is provided in Grosz (1974a, 1974b).
[40] A point that Grice (1975, p. 48) himself makes.

to provide evidence that their doer is engaged in something that perceivers find understandable, even if they are not favored thereby. This paper has dealt only with utterances, but some of what it has dealt with about utterances could almost as well be dealt with about entirely nonlinguistic doings.

In sum, then, whenever we come into contact with another through the mails, over the telephone, in face-to-face talk, or even merely through immediate co-presence, we find ourselves with one central obligation: to render our behavior understandably relevant to what the other can come to perceive is going on. Whatever else, our activity must be addressed to the other's mind, that is, to the other's capacity to read our words and actions for evidence of our feelings, thoughts, and intent. This confines what we say and do, but it also allows us to bring to bear all of the world to which the other can catch allusions.

REFERENCES

Austin, J. L. 1965. *How to Do Things with Words.* Edited by J. O. Urmson. New York: Oxford University Press.

Baker, Charlotte. 1975. "This Is Just a First Approximation, But. . . ." *Chicago Linguistic Society* 11:37–47.

Berger, Peter L., and Hansfried Kellner. 1964. "Marriage and the Construction of Reality." *Diogenes* 45:1–25.

Bloomfield, Leonard. 1933. "Substitution." Pp. 247–63 in *Language,* by Leonard Bloomfield. New York: Holt.

Chafe, Wallace L. 1974. "Language and Consciousness." *Language* 59:112.

———. 1976. "Givenness, Contrastiveness, Definiteness, Subjects and Topics." Pp. 27–55 in *Subject and Topic,* edited by Charles N. Li. New York: Academic Press.

Clark, Herbert H. 1979. "Responding to Indirect Speech Acts." *Cognitive Psychology* 11:4.

Clark, Herbert H., and Catherine Marshall. 1981. "Definite Reference and Mutual Knowledge." Pp. 10–63 in *Elements of Discourse Understanding,* edited by A. Joshi, B. Webber, and I. Sag. Cambridge: Cambridge University Press.

Coulter, Jeff. 1979. "Beliefs and Practical Understanding." Pp. 163–86 in *Everyday Language: Studies in Ethnomethodology,* edited by George Psathas. New York: Irvington.

Derrida, Jacques. 1977. "Limited Inc abc . . ." *Glyph* 2:162–254.

Gardner, Carol. 1983. "Aspects of Gender Behavior in a Small Southwestern City." Ph.D. dissertation, Department of Sociology, University of Pennsylvania.

Garfinkel, Harold. 1967. "Studies of the Routine Grounds of Everyday Activities." Pp. 35–75 in *Studies in Ethnomethodology,* by Harold Garfinkel. Englewood Cliffs, N.J.: Prentice-Hall.

Garnes, Sara, and Zinny S. Bond. 1975. "Slips of the Ear: Errors in Perception of Casual Speech." *Chicago Linguistic Society* 11:214–25.

Garvey, Catherine. 1975. "Requests and Responses in Children's Speech." *Journal of Child Language* 2:41–63.

Goffman, Erving. 1971. *Relations in Public.* New York: Harper Colophon.

———. 1981. *Forms of Talk.* Philadelphia: University of Pennsylvania Press.

Grice, H. Paul. 1975. "Logic and Conversation." Pp. 41–58 in *Syntax and Semantics.* Vol. 3, *Speech Acts,* edited by Peter Cole and Jerry L. Morgan. New York: Academic Press.

Grimshaw, Allen D. 1980. "Mishearings, Misunderstandings, and Other Nonsuccesses in Talk." *Sociological Inquiry* 50:31–74.

Grosz, Barbara J. 1974a. "The Representation and Use of Focus in Dialogue Under-

standing." Stanford Research Institute Project 1526, Technical Note 90. Menlo Park, Calif.: SRI.

————. 1974b. "The Structure of Task Oriented Dialogues." Stanford Research Institute Project 1526, Technical Note 90. Menlo Park, Calif.: SRI.

Gumperz, John, and Hannah Kaltman. 1980. "Prosody, Linguistic Diffusion and Conversational Inference." Pp. 44–65 in *Proceedings of the Berkeley Linguistics Society*, edited by Bruce R. Caron et al. Berkeley: Berkeley Linguistics Society, University of California.

Halliday, M. A. K. 1967. "Notes on Transitivity and Theme in English, Part II." *Journal of Linguistics* 20:199–244.

Halliday, M. A. K., and R. Hasan. 1976. *Cohesion in English*. London: Longmans.

Jefferson, Gail. 1978. "Sequential Aspects of Storytelling in Conversation." Pp. 219–48 in *Studies in the Organization of Conversational Interaction*, edited by Jim Schenkein. New York: Academic Press.

Keenan, Elinor Ochs, and Bambi Schieffelin. 1976a. "Foregrounding Referents: A Reconsideration of Left Dislocation in Discourse." Pp. 240–57 in *Proceedings of the Berkeley Linguistics Society*, edited by Bruce R. Caron et al. Berkeley: Berkeley Linguistics Society, University of California.

————. 1976b. "Topic as a Discourse Notion: A Study of Topic in the Conversations of Children and Adults." Pp. 335–84 in *Subject and Topic*, edited by Charles N. Li. New York: Academic Press.

Kiparsky, Paul, and Carol Kiparsky. 1970. "Fact." Pp. 143–73 in *Progress in Linguistics*, edited by Manfred Bierwisch and Karl Erick Heidolph. Paris: Mouton.

Labov, William. 1972. "The Transformation of Experience in Narrative Syntax." Pp. 354–96 in *Language in the Inner City*, by William Labov. Philadelphia: University of Pennsylvania Press.

Lakoff, Robin. 1971. "If's, And's, and But's about Conjunctions." Pp. 115–50 in *Studies in Linguistic Semantics*, edited by Charles Fillmore and D. T. Langendoen. New York: Holt.

————. 1973a. "The Logic of Politeness; or, Minding Your P's and Q's." Pp. 292–305 in *Papers from the 9th Regional Meeting, Chicago Linguistic Society*, edited by Claudia Corum et al. Chicago: Chicago Linguistic Society.

————. 1973b. "Questionable Answers and Answerable Questions." Pp. 453–67 in *Issues in Linguistics*, edited by Braj B. Bachru et al. Urbana: University of Illinois Press.

Lehman, Christina. 1977. "A Re-Analysis of Givenness: Stress in Discourse." *Chicago Linguistic Society* 13:316–24.

McTear, Michael F. 1977. "Starting to Talk: How Pre-School Children Initiate Conversational Exchanges." Belfast Working Papers in Language and Linguistics no. 2. Belfast: Northern Ireland Polytechnic.

Maynard, Douglas W. 1980. "Placement of Topic Changes in Conversation." *Semiotica* 30:263–90.

Merritt, Marilyn. 1976. "On Questions Following Questions in Service Encounters." *Language in Society* 5:315–57.

Nelson, Katherine, and Janice Gruende. 1979. "At Morning It's Lunchtime: A Scriptal View of Children's Dialogues." *Discourse Processes* 2:73–94.

Nofsinger, Robert E., Jr. 1975. "The Demand Ticket: A Conversational Device for Getting the Floor." *Speech Monographs* 42:1–9.

Nunberg, Geoffrey. 1979. "The Non-Uniqueness of Semantic Solutions: Polysemy." *Linguistics and Philosophy* 3:143–84.

Prince, Ellen F. 1979. "On the Given/New Distinction." *Chicago Linguistic Society* 15:267–78.

————. 1981. "Toward a Taxonomy of Given/New Information." Pp. 223–55 in *Radical Pragmatics*, edited by Peter Cole. New York: Academic Press.

Psathas, George, and Martin Kogloff. 1976. "The Structure of Directions." *Semiotica* 17:111–30.

Sacks, Harvey. 1972a. "An Initial Investigation of the Usability of Conversational Data for Doing Sociology." Pp. 31–74 in *Studies in Social Interaction*, edited by David Sudnow. New York: Free Press.

————. 1972b. "On the Analyzability of Stories by Children." Pp. 329–45 in *Directions in Sociolinguistics,* edited by John J. Gumperz and Dell H. Hymes. New York: Holt, Rinehart & Winston.

Sacks, Harvey, and Emanuel A. Schegloff. 1979. "Two Preferences in the Organization of Reference to Persons in Conversation and Their Interaction." Pp. 15–21 in *Everyday Language: Studies in Ethnomethodology,* edited by George Psathas. New York: Irvington.

Sacks, Harvey, Emanuel A. Schegloff, and Gail Jefferson. 1974. "A Simplest Systematics for the Organization of Turn-Taking for Conversation." *Language* 50:696–735.

Sag, Ivan A., and Mark Liberman. 1975. "The Intonational Disambiguation of Indirect Speech Acts." *Chicago Linguistic Society* 11:487–97.

Sag, Ivan A., and Ellen F. Prince. 1979. "Bibliography of Works Dealing with Presuppositions." Pp. 389–403 in *Syntax and Semantics.* Vol. 11, *Presupposition,* edited by C.-K. Oh and D. Dineen. New York: Academic Press.

Schank, Roger. 1977. "Rules and Topics in Conversation." *Cognitive Science* 1:421–41.

Schank, Roger, and Robert P. Abelson. 1977. "Scripts, Plans, Goals, and Understanding." Hillsdale, N.J.: Erlbaum.

Schegloff, Emanuel A. 1968. "Sequencing in Conversational Openings." *American Anthropologist* 70:1075–95.

————. 1972. "Notes on a Conversational Practice: Formulating Place." Pp. 75–119 in *Studies in Social Interaction,* edited by David Sudnow. New York: Free Press.

Schegloff, Emanuel A., and Harvey Sacks. 1973. "Opening Up Closings." *Semiotica* 8: 289–327.

Schiffrin, Deborah. 1982. "Discourse Markers: Semantic Resource for the Construction of Conversation." Ph.D. dissertation, Department of Linguistics, University of Pennsylvania.

Searle, John R. 1969. *Speech Acts: An Essay in the Philosophy of Language.* Cambridge: Cambridge University Press.

Sherzer, Joel. 1973. "On Linguistic Semantics and Linguistic Subdisciplines: A Review Article." *Language in Society* 2:269–89.

Shugar, Grace. 1978. "A Discourse Analysis System Applied to Talk of Children at Age Three to Five." Paper presented at the Discourse Symposium, Fifth International Congress of Applied Linguistics, Montreal, August 21–26.

Snow, Catherine E. 1977. "The Development of Conversation between Mothers and Babies." *Journal of Child Language* 4:1–22.

Stalnaker, Robert. 1973. "Presuppositions." *Journal of Philosophical Logic* 2:447–57.

Stenning, Keith. 1978. "Anaphora as an Approach to Pragmatics." Pp. 162–200 in *Linguistic Theory and Psychological Reality,* edited by J. Bresnan, M. Halle, and G. Miller. Cambridge, Mass.: MIT Press.

Tasmowsky-De Ryck, Liliane, and S. Paul Verluyten. 1981. "Pragmatically Controlled Anaphora and Linguistic Form." *Linguistic Inquiry* 12:153–54.

Weiser, Ann. 1975. "How Not to Answer a Question: Purposive Devices in Conversational Strategy." *Chicago Linguistic Society* 11:649–60.

Structural Equations and Path Analysis for Discrete Data[1]

Christopher Winship
Northwestern University and *National Opinion Research Center*

Robert D. Mare
University of Wisconsin—Madison

This article proposes a solution to the long-standing methodological problem of incorporating discrete variables into causal models of social phenomena. Only a subset of the variety of ways in which discrete data arise in empirical social research can be satisfactorily modeled by conventional log-linear or logit approaches. Drawing on the insights of several literatures, this article exposits a general approach to causal models in which some or all variables are discretely measured and shows that path analytic methods are available which permit quantification of causal relationships among variables with the same flexibility and power of interpretation as is feasible in models that include only continuous variables. It presents methods of identifying and estimating these models and shows how the direct and indirect effects of independent variables can be calculated by extensions of usual path analysis methods for continuous variables. An important distinction developed here is that discrete variables can play two roles: (1) as measures of inherently discrete phenomena and (2) as indicators of underlying continuous variables. The value of this distinction is shown in two empirical examples examined previously by other authors. In examining the effects of social background and parental encouragement on college plans of high school seniors, the article shows that modeling a discrete measure of encouragement as an indicator of a latent continuous variable rather than as an inherently discrete variable (as has been done in previous analyses) provides a clearer interpretation and a superior fit to the data. In examining the effects of state Fair-Employment-Practices Legislation on black-white wage differentials, this study shows that two distinct effects on the relative wage can be detected: the direct ameliorative effect of the law itself and the effect of the popular progressive sentiment for racial equality of which the law is an indicator. The methods and models presented here are not only natural generalizations of structural equation and path analysis methods for continuous variables to include discrete variables but also provide a means of investigating a richer variety of substantive hypotheses than is feasible with methods for discrete data commonly used in the sociological literature to date.

One of the most important methodological developments in the social sciences in recent decades has been the use of structural equation models to specify and interpret causal relationships among variables (Goldberger and Duncan 1973; Duncan 1975; Bielby and Hauser 1977). Following upon related developments in other disciplines (Goldberger 1971), structural equation models in sociology have been applied primarily to processes in which the dependent variables are continuously scaled. Thus these models have been amenable to empirical analyses that use the general linear model and its extensions. More recently, causal analysis has been applied to discrete response variables (Goodman 1972, 1973a, 1973b, 1979; Fienberg 1980). Goodman and others have shown how to specify causal relationships among discrete variables and quantify the relationships by application of log-linear or logit models. These methods permit analysts to impose causal interpretations on multivariate systems of discrete variables; illuminating applications now appear in the literature (e.g., Gortmaker 1979).

At the same time, however, commentators have pointed out that causal models constructed from log-linear and logit models have limitations and that these models are not directly analogous to causal models with continuous variables (Fienberg 1975, 1980; Rosenthal 1980). Although there has been considerable interest in developing both a path analysis for discrete variables and one based on log-linear and logit models (e.g., Leik [1975], and the series of articles in *Sociological Methodology 1976* on this subject [Heise 1975]), no one has yet presented a satisfactory solution to this problem.[2] Indeed, several commentators have suggested that such a path analysis cannot be developed (Davis and Schooler 1974; Davis 1975; Fienberg 1980).

The limitations of the log-linear and logit models derive in part from the nonlinear functional forms of relationships among discrete variables that are implicit in systems of logit equations. In these models, discrete endogenous variables are not subject to consistent treatment. In partic-

[2] Davis (1975) proposes a path analysis for discrete data based on the linear probability model. Unlike logit models and the other approaches discussed here, however, this approach assumes that the effects of independent variables are constant over the entire (0,1) range of the dependent variable. The shortcomings of this model are well-known (Hanushek and Jackson 1977, pp. 183–86).

ular, when a discrete intervening variable is the dependent variable, it appears as the logit of a probability; when it is an independent variable, it appears as a dummy variable. For example, consider a simple model in which social background factors affect whether or not an individual has a college degree, and background and degree status affect whether or not the individual experiences unemployment. As a dependent variable, degree status is measured as the logit transformed probability of having a degree; as an independent variable, it is a simple dummy variable. Although suitable in many applications, this treatment of discrete variables may not be suitable in others. In some applications it is, at best, an arbitrary formulation dictated by the use of log-linear models and not by its reasonableness in substantive applications.

Discrete variables arise in a variety of contexts. Some measure inherently discrete characteristics, for example, sex or race. Some arise out of imperfections in measurement instruments. For example, family income, a continuous variable, may be measured in survey data as the discrete variable, "less than or greater than $20,000 per year." Some measure inherently discrete observations that are of interest less in their own right than as indicators of latent continua. For example, whether or not a white respondent would be willing to live in a predominantly black neighborhood is a binary response that derives its substantive interest from its indication of underlying racial tolerance. Finally, some discrete variables may play dual roles; that is, they can act both as direct measurements of discrete phenomena and as indicators of unmeasured phenomena that are best conceived of as continuous. For example, whether or not a state or nation has a particular type of law is clearly discrete, but this fact may index the accumulated sentiments of populations or of decision-making bodies. In some analyses both the law itself and the sentiments it reflects may be of interest.

These alternative contexts in which discrete variables arise require analysts to select statistical models that are in accord with the kinds of effects implied by substantive reasoning. In many contexts, recently developed methods for the causal analysis of discrete data embody the substantive ideas of interest satisfactorily. In others, however, alternative models may be more appropriate.

An issue related to the proper formulation of causal models for discrete data is the capacity of such models to allow the application of path analysis procedures to discrete data. Structural equation models for continuous data allow the analyst to quantify the direct and indirect effects of predetermined variables on endogenous variables. In discrete systems, analyzed by existing methods, however, the inconsistent treatment of intervening variables implies that the usual "theorem of path analysis" (Duncan 1966) cannot be applied directly. This has led some commen-

tators to conclude that "there is no calculus of paths" for these models (Fienberg 1980, p. 120).

This article presents several alternative approaches to the formulation of structural equation models for discrete data that take into account the variety of roles played by discrete variables in quantitative analysis. It shows that the current approach to causal analysis of discrete data is one of several that may be applied, depending on the underlying causal process that the analyst hypothesizes. In addition, the article shows that the difficulties encountered in applying path analytic principles have attractive solutions regardless of which approach to the analysis of discrete systems is adopted. It also shows that structural equation models for discrete data, though sometimes computationally complex, are every bit as flexible analytically as corresponding models for continuous variables. Indeed, with little difficulty, discrete and continuous response variables can be analyzed together in the models discussed here.

The insights from several literatures are synthesized in this study. There is an examination of alternative formulations of discrete variables when they are dependent variables; that is, either as the outcomes of inherently discrete stochastic processes (Yule 1900) or as realizations of latent continuous variables (Pearson 1900).[3] There is also consideration, based mostly on recent work by Heckman (1978), of the alternative formulations of the role of discrete variables when they are independent variables; that is, either as observed dummy variables or as indicators of unobserved continuous variables. In incorporating these alternative conceptions into structural equation models, the article applies the recent literature on simultaneous probit models (Zellner and Lee 1965; Ashford and Sowden 1970; Heckman 1978; Amemiya 1974, 1976, 1978; Muthén 1979) and on discrete choice (Hausman and Wise 1978; Manski 1981; McFadden 1976, 1980). Finally, it is shown that the problem of calculating direct and indirect effects in structural equation models can be solved by straightforward application of methods proposed by Stolzenberg (1979) for the analysis of nonlinear models. The contribution of this article, therefore, is to apply the contributions of diverse literatures to a central methodological problem in sociology.

This study is limited in two respects. First, only recursive structural equation models are considered. The ideas presented here can be extended to systems embodying simultaneity, but the latter are beyond the scope of this article. Discussions of nonrecursive systems are available elsewhere (Heckman 1976, 1978; Brier 1978; Zellner and Lee 1965). In any case, it is from recursive models that structural equation models have thus far borne their greatest fruit in sociology (e.g., Duncan, Featherman, and

[3] The possibility that discrete variables are indicators of latent categorical variables ("latent classes") is not considered here (Lazarsfeld and Henry 1968; Goodman 1974; Clogg 1980).

Duncan 1972; Sewell and Hauser 1975). Second, the treatment of discrete variables presented here is limited to variables with only two categories. Models with ordered response variables can, however, be developed from the ideas for dichotomous responses presented here (e.g., Amemiya 1981; McKelvey and Zavoina 1975).

The balance of the article is divided into four major sections. Section I discusses the problem of conceptualizing discrete variables when they are embedded in structural equation models. First it discusses discrete variables both as dependent variables and as independent variables, then it presents four alternative recursive models that embody distinct relationships between a discrete variable and other variables in the model. Section II discusses the identification of recursive models with discrete data. Section III shows how the methods of path analysis can be adapted to models with discrete variables. Section IV presents two numerical examples that illustrate the methods exposited in Sections I, II, and III. Methods of estimating the models discussed in the article are discussed in the Appendix.

I. THE CONCEPTUALIZATION OF DISCRETE VARIABLES

This section presents alternative formulations for discrete variables, first as dependent variables and then as independent variables. Then it applies these formulations to four simple recursive models containing discrete variables.

1. Discrete Variables as Dependent Variables

Historically, there have been two approaches to modeling discrete dependent variables: they have been modeled as observed indicators of unobserved continuous variables and as inherently discrete outcomes of binomial trials.

Discrete variables as realizations of underlying continuous variables: threshold model.—Discrete variables may be thought of as indicators of continuous variables that are either difficult or impossible to measure directly. Let d_y be a dichotomous variable with the value 1 or 0 and let Y^* be an unobserved continuous variable. Here and throughout the article, unobserved variables such as Y^* will be assumed to have means of zero and unit variance. Alternative scaling assumptions do not materially affect the models and methods presented. Under this model, Y^* and d_y are related as follows:

$$d_y = 1 \text{ if } Y^* \geq L ,$$
$$d_y = 0 \text{ if } Y^* < L ; \tag{1}$$

where L is a parameter denoting the "threshold" across which d_y changes from 0 to 1. Thus Y^* and d_y are related by a nonlinear transformation: all values of Y^* above or equal to L have been transformed to 1; all values below L have been transformed to 0. There are numerous instances where this formulation is applicable. An attitude item on whether or not the respondent is willing to live in a predominantly black neighborhood is a binary outcome that may index whether the respondent is above or below a threshold on an unmeasurable continuum of racial tolerance/prejudice. Data obtained on whether or not an individual has attended college may index the approximately continuous variable, length of time in school, which was not obtained in a survey or published tabulation. If Y^* is affected by an independent variable, say X, then a model may be written as

$$Y^* = \beta_0 + \beta_1 X + \epsilon_y , \qquad (2)$$

where β_0 and β_1 are parameters to be estimated, ϵ_y is a stochastic disturbance assumed to be uncorrelated with X, and Y^* is related to the observed binary variable as in (1) above.[4] The disturbance ϵ_y is assumed to follow a probability distribution that remains to be specified. If ϵ_y has a normal distribution, (1) and (2) define a probit model (Hanushek and Jackson 1977, pp. 204–5). If ϵ_y has an extreme value distribution (Johnson and Kotz 1970, pp. 272–89), (1) and (2) define a logit model (McFadden 1974). The logit and probit models are discussed in more detail in the Appendix.

Figures 1A and B show graphically the essential features of this model for discrete response variables. Figure 1A shows that the dichotomous variable is an exact step function of Y^*. Figure 1B provides a path diagram of the model. The wavy line connecting Y^* and d_y denotes that the relationship is nonlinear and deterministic; the solid line connecting X and Y^* denotes a linear, stochastic relationship.

Discrete variables as outcomes of binomial trials.—A second model for binary data is that observed dichotomous variables arise from binomial trials; that is, that the phenomena of interest are inherently discrete and observed binary outcomes are stochastic in nature. This can be written

---

[4] Here and elsewhere models are presented with a single independent variable X. Adding additional independent variables to the models raises no new issues beyond those discussed for single independent variables. An alternative equivalent specification of the threshold model is to regard the threshold L as random, rather than fixed, and the latent continuous variable Y^* as a deterministic rather than stochastic function of X. That is, $L = \bar{L} + \epsilon_L$, and $Y^* = \beta_0 + \beta_1 X$; where L is now a random threshold, $\bar{L}$ is the mean of L, and ϵ_L is a stochastic disturbance. If $\epsilon_L = \epsilon_y$, this model is equivalent to the one discussed above. Although the model presented here postulates a stochastic relationship between d_y and Y^*, it differs from the stochastic model presented below in which Y^* is not only stochastically related to d_y but also a function of X and a stochastic disturbance (see n. 7 below). To simplify the exposition, the random threshold formulation is not adopted in the text.

X affects the binary outcome, a model can be specified to allow X to affect the probability that the outcome is a 1 (or a 0). That is,

$$p = F(\beta_0 + \beta_1 X) , \qquad (4)$$

where F is the cumulative density (distribution) function for some symmetric probability distribution that must be specified and β_0 and β_1 are parameters.[5] If F is the logistic function, (4) is a logit model.[6] If F is the cumulative normal density function, (4) is a probit model. An alternative, but equivalent, way of expressing the effect of X on the discrete outcome is to define an unobserved variable Y^* which is a transformation of the unobserved probability p such that it is a linear function of X and an error term ϵ_y. That is,

$$Y^* = \beta_0 + \beta_1 X + \epsilon_y , \qquad (5)$$

with

$$
\begin{aligned}
d_y &= 1 \text{ with probability } G(Y^*) , \\
d_y &= 0 \text{ with probability } 1 - G(Y^*) ,
\end{aligned}
\qquad (6)
$$

where G is the cumulative density function of a distribution that remains to be specified. Then the standard binomial trials model is obtained by assuming that ϵ_y is zero for all observations, that is, that all observations with the same X have equivalent probabilities that $d_y = 1$ (Amemiya and Nold 1975; Hanushek and Jackson 1977, pp. 99–200, 203). As is demonstrated below, without this assumption the model is not identified.

Figures 2A and B represent the essential features of the binomial trials model as formulated by equations (5) and (6). Figure 2A shows the relationship between the transformation of the unobserved probability and the observed dummy variable. In contrast to the threshold model where the continuous and discrete variables are related by a nonlinear but deterministic function, in the binomial trials model, Y^* and d_y are nonlinearly and stochastically related. Figure 2B presents a path diagram of the effect of X on d_y. The dotted line denotes the nonlinear stochastic relationship plotted in figure 2A.

Relationship between the two models.—The binomial trials and thresh-

---

[5] For probability distributions of general shape,

$$p = \int_{-(\beta_0 + \beta_1 X)}^{\infty} f(u)\,du = 1 - F[-(\beta_0 + \beta_1 X)] ,$$

where f is a probability density function and F is its corresponding distribution function. For symmetric F, this expression simplifies to (4).

[6] The logistic function is the cumulative density function of the extreme value distribution (Johnson and Kotz 1970, pp. 272–89).

old models are related as follows. Let Y' be a continuous unobserved variable with mean 0 and variance 1 such that

$$d_y = 1 \text{ if } Y' \geq M ,$$

$$d_y = 0 \text{ if } Y' < M . \tag{7}$$

Then the binomial trials model may be written

$$Y^* = \beta_0 + \beta_1 X + \epsilon_y , \tag{8}$$

$$Y' = \lambda Y^* + \epsilon_{y'} , \tag{9}$$

where λ is the zero-order coefficient for the regression of Y' on Y^* and $\epsilon_{y'}$ is a stochastic disturbance. In this formulation, there is a variable Y' which has a threshold M dividing observations with $d_y = 1$ from those
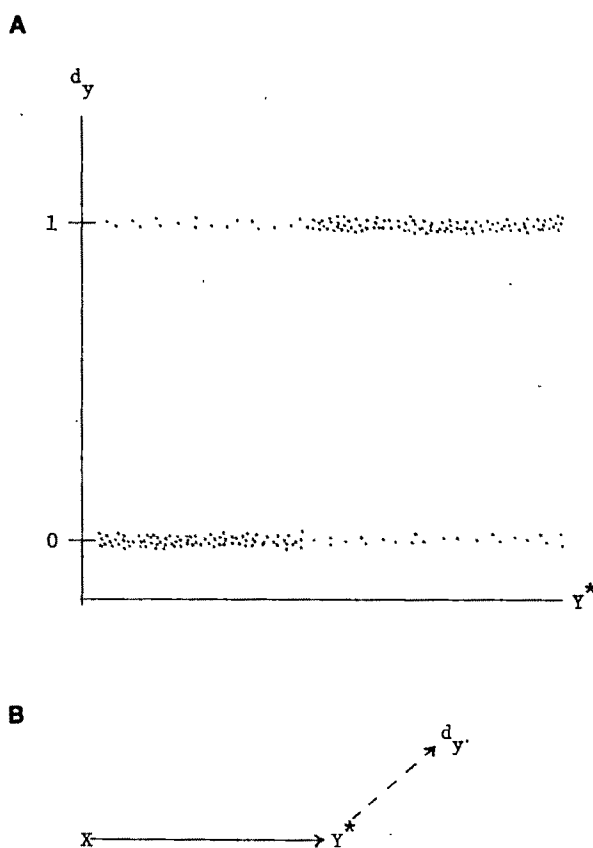

FIG. 2.—Binomial trials model. A, Relationship between d_y and Y^*. B, Path diagram.

with $d_y = 0$. The variable Y', however, is not itself a direct linear function of the independent variable X but instead is affected linearly by Y^* which, as before, is a transformation of the unobserved $p(d_y = 1)$ and is a stochastic linear function of X. Equations (8) and (9) are a structural equation and a measurement equation, respectively. Thus d_y can be interpreted as a discrete variable with measurement error, which imperfectly classifies observations into whether Y^* is less or greater than M. The threshold model is a special case of equations (7), (8), and (9) where $\lambda = 1$ and $\epsilon_{y'} = 0$. Thus the threshold model is a restricted case of the binomial trials model that assumes no measurement error in d_y. In practice, it may make no difference whether or not λ can be separated from the structural relationship between X and Y^*.[7]

In the past, the relative merits of the threshold and binomial models of discrete variables have occasioned considerable debate, notably between Yule, who advocated the binomial trials approach, and Pearson, who advocated the threshold approach (Fienberg 1980). In recent decades, this debate has translated into the parallel lines of intellectual development of log-linear models for discrete data, a method which has assumed a binomial trials model, and biometric and econometric approaches (Ashford and Sowden 1970; Zellner and Lee 1965) which have assumed a threshold model. As will be shown below, each approach has considerable applicability to the analysis of discrete data in the social sciences depending on the substantive context and the purposes of the analyst (Goodman 1981).

2. Discrete Variables as Independent Variables

The preceding discussion showed that discrete variables may arise either as indicators of unobserved continuous variables that may be of greater substantive interest or as results of unobserved (continuous) probabilities (or their transformations) that lack substantive content. In both instances, however, they arise as continuous variables that are linearly related to independent variables. In contrast, when discrete variables are treated as independent variables, they may occupy two distinct roles. That is, they may measure discrete phenomena of direct substantive interest or they may be indicators of unobserved continuous variables.[8]

[7] An alternative way to relate these two models is to assume that, in the threshold model, the thresholds are random, rather than fixed; i.e., $L = \bar{L} + \epsilon_L$ (see n. 4 above). This model can be shown to be equivalent to the binomial trials model as follows: substitute (9) into (7) to obtain $d_y = 1$ if $Y^* \geq (M - \epsilon_{y'})/\lambda$, $d_y = 0$ if $Y^* < (M - \epsilon_{y'})/\lambda$. Then the two models are equivalent if $\bar{L} = M/\lambda$ and $\epsilon_L = \epsilon_{y'}/\lambda$. Note that this model generalizes the random threshold model discussed in n. 4 above by relaxing the assumption made in the latter model that Y^* is an exact linear function of X.

[8] This discussion draws much from Heckman (1978).

Discrete variables as indicators of discrete phenomena.—A discrete variable may have a direct effect on a dependent variable. In this case it enters a linear equation as a dummy variable denoting that the group or characteristic of interest has a different intercept from other observations in the analysis, that is, there is a "structural shift" associated with the group or characteristic (Heckman 1976, 1978). If Z is an endogenous variable (which may be either an observed continuous variable or an unobserved continuous variable with discrete indicators), this type of effect of a discrete variable is denoted by α_2 in the equation $Z = \alpha_0 + \alpha_1 X + \alpha_2 d_y + \epsilon_z$, where α_0, α_1, and α_2 are parameters, X is an independent variable, ϵ_z is a stochastic disturbance, and, as before, d_y is a dummy variable taking the value 1 or 0.

Discrete variables as indicators of unobserved continuous variables.— Alternatively, an observed discrete variable itself may exert no effect on a dependent variable but instead may be an indicator of an unobserved continuous variable that affects the dependent variable. Such a model can be written $Z = \alpha_0 + \alpha_1 X + \alpha_3 Y^* + \epsilon_z$, where $d_y = 1$ with probability $F(Y^*)$, $d_y = 0$ with probability $1 - F(Y^*)$, and all notation is as defined above.

These two models reflect distinct conceptions of how a discrete variable exerts its effect. In the first case, the effect of the discrete variable is direct; in the second, the observed variable is merely an indicator for the continuous variable of interest. Consider several examples of this distinction.

1. Political party membership, indicated by a discrete variable (Republican versus Democrat), may affect an individual's attitude toward proposed legislation. If the legislation is proposed by an individual's own party and party leadership is followed, the discrete variable has a direct positive effect on the individual's attitude. Alternatively, an individual may be a party member but support legislation because of her or his ideological beliefs. In this case, there is no structural relationship between membership and attitude toward legislation. However, a continuous, unobserved variable, ideology, of which party membership *may* be an indicator, might in itself affect attitude toward legislation.

2. Marital disruption may affect the chances that the children of separated spouses will drop out of high school. Whether or not parents have separated (a discrete variable) may affect dropout chances. Alternatively, the separation of parents may index family conflict or disharmony (an unobserved variable), which may exert more of an effect on the probability of dropping out. These alternative conceptions imply distinct models for the discrete variable, marital disruption. Indeed, they harbor distinct implications for the desirability of parents in conflict-ridden marriages staying together until their children have completed school.

3. Whether or not an infant has a low birth weight has a pronounced effect on its chances of survival (Gortmaker 1979). Measuring "low birth weight" as a dichotomy, low weight or not low weight, implies that there is a critical weight below which an infant's chances of death increase astronomically. It may, however, be more reasonable to assume that the relationship between birth weight and mortality is continuous (albeit non-linear) and that the dichotomous measure should be regarded as an indicator of a continuous measure of weight.

These examples illustrate that the way in which discrete variables should be treated as independent variables in structural equation models is closely linked to substantive arguments about the way their effects are hypothesized to occur.

3. Some Recursive Models

The foregoing ideas can be combined to yield recursive models which reflect alternative assumptions about the role of discrete variables. Four models are considered here. In each a single, predetermined variable X, which may be either discrete or continuous, affects an intervening variable Y, which is observed as a discrete binary variable d_y, and both X and Y affect a third variable Z, which may be discrete or continuous. These models can be extended to include additional predetermined and intervening variables. The models are presented here and their identifiability is discussed in the next section.

Model I: intervening variable as unobserved continuous variable deterministically related to observed discrete variable.—Consider first the model represented by the path diagram in figure 3. The model may be written

$$Z = \alpha_0 + \alpha_1 X + \alpha_2 Y^* + \epsilon_z ,$$

$$Y^* = \beta_0 + \beta_1 X + \epsilon_y ;$$

where

$$\text{cov} (\epsilon_z, \epsilon_y) = 0 ,$$

$$d_y = 1 \text{ if } Y^* \geq L ,$$

$$d_y = 0 \text{ if } Y^* < L ,$$

and all notation is as defined above. In this model the observed discrete variable d_y does not enter the two structural equations of the model. Instead it is an indicator of an unobserved variable Y^*, to which it is deterministically related as in figure 1A. As will be shown below, it is

65

possible to identify and estimate the structural parameters α_0, α_1, α_2, β_0, and β_1.

Model II: intervening variable as unobserved continuous variable stochastically related to observed discrete variable.—Figure 4A represents model II in which the discrete variable d_y is an indicator for an unobserved continuous variable as in model I. Here, however, d_y is the outcome of binomial trials, and thus Y^* and d_y are related stochastically as in figure 2A. The model may be written

$$Z = \alpha_0 + \alpha_1 X + \alpha_2 Y^* + \epsilon_z ,$$

$$Y^* = \beta_0 + \beta_1 X + \epsilon_y ;$$

where

$$\text{cov } (\epsilon_z, \epsilon_y) = 0 ,$$

$$p(d_y = 1) = F(Y^*) ,$$

$$p(d_y = 0) = 1 - F(Y^*) ,$$

and, as above, F is a cumulative probability function such as the logistic or cumulative normal.

An equivalent, alternative formulation of model II can be obtained by defining a continuous unobserved variable Y' which is linearly (though not exactly) related to Y^* and is nonlinearly but deterministically related
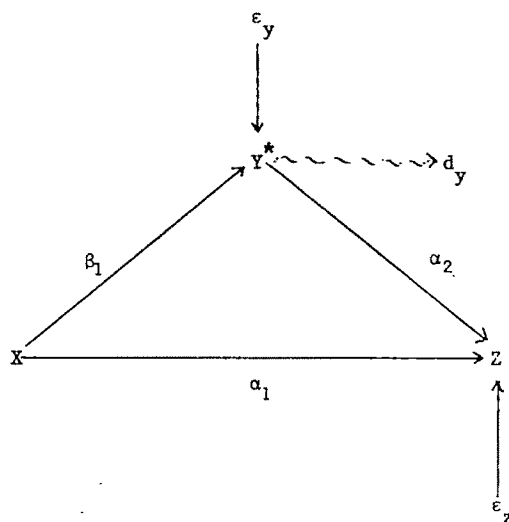
FIG. 3.—Path diagram of model I

to d_y. Thus, as in the preceding section, Y' is a variable which is classified perfectly by d_y, just as Y^* is classified perfectly by d_y in model I. Model II may then be represented as in figure 4B, which shows that Y^* and Y' are related by the measurement equation $Y' = \lambda Y^* + \epsilon_{y'}$, where $d_y = 1$ if $Y' \geq M$, $d_y = 0$ if $Y' < M$. This formulation shows that model II contains one more parameter (λ) than model I, indicating that the relationship between the unobserved continuous variable and its discrete indicator d_y is not exact and needs to be estimated from the data. As will be shown below, the parameters of model II are not identifiable, in gen-
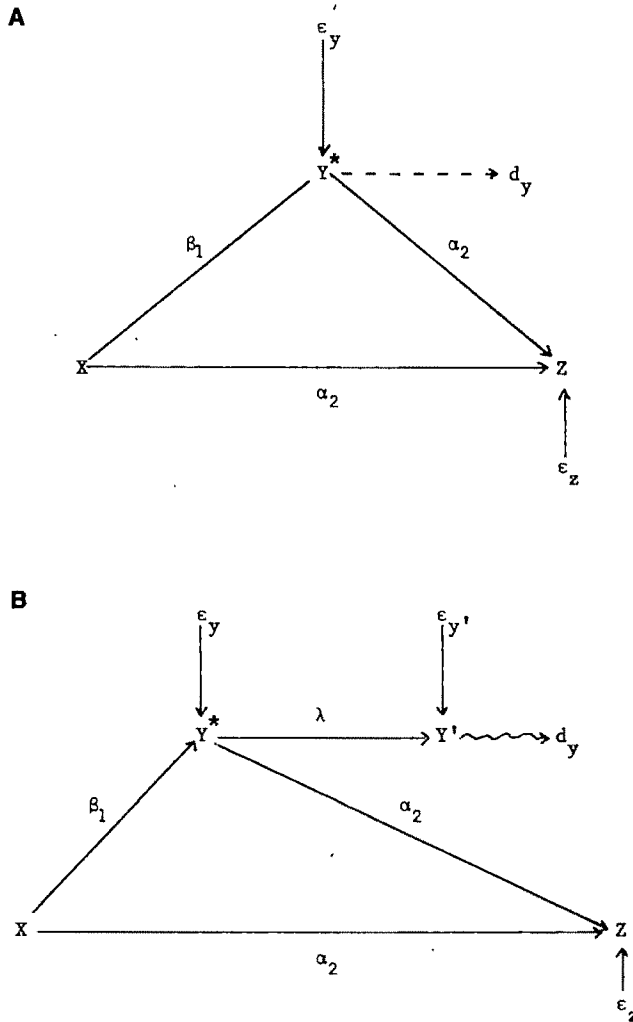
FIG. 4.—Model II. *A*, Path diagram of model II. *B*, Path diagram of model II: threshold representation.

67

eral, without additional information. This information, however, is often available, and simple modifications of model II are identifiable.

Model III: intervening variable as observed discrete variable deterministically related to unobserved continuous variable.—In contrast to the two models discussed above, in the third and fourth models the observed discrete intervening variable d_y affects Z directly. In model III, d_y and an unobserved continuous variable Y^*, which is a linear function of X, are related deterministically; thus this model parallels model I. Model III is diagramed in figure 5. The model is

$$Z = \alpha_0 + \alpha_1 X + \alpha_3 d_y + \epsilon_z \,,$$

$$Y^* = \beta_0 + \beta_1 X + \epsilon_y \,;$$

where

$$\text{cov } (\epsilon_z, \epsilon_y) = 0 \,,$$

$$d_y = 1 \text{ if } Y^* \geq L \,,$$

$$d_y = 0 \text{ if } Y^* < L \,.$$

Figure 5 indicates that model III differs from typical three-variable recursive models because it contains an unobserved variable Y^* which is related nonlinearly to an observed variable d_y but (in contrast to models I and II) both variables enter the structural equations. The structural
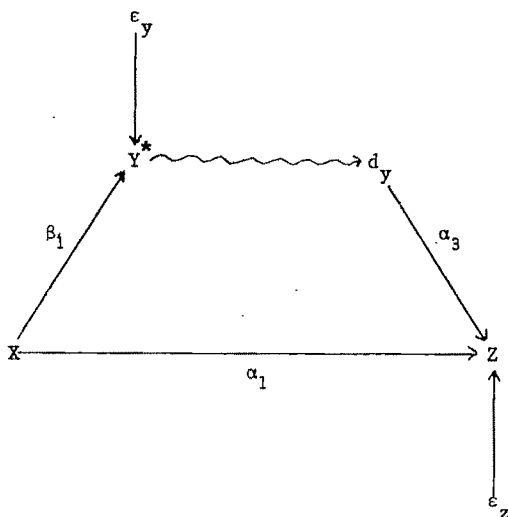
FIG. 5.—Path diagram of model III

parameters α_0, α_1, α_3, β_0, and β_1 are nonetheless identifiable, as will be shown in the next section.

Model IV: intervening variable as observed discrete variable stochastically related to unobserved continuous variable.—In model IV, d_y and an unobserved continuous variable Y^*, which is a linear function of X, are related stochastically. Thus this model parallels model II. That is, d_y is the outcome of binomial trials. Model IV is diagramed in figure 6A
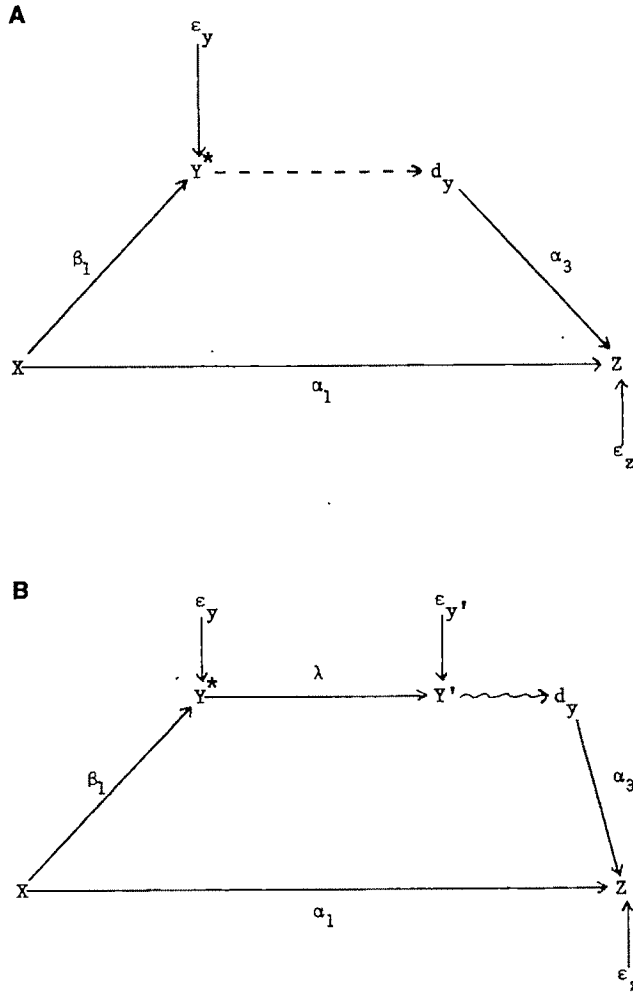
Fig. 6.—Model IV. *A*, Path diagram of model IV. *B*, Path diagram of model IV: threshold representation.

and may be written

$$Z = \alpha_0 + \alpha_1 X + \alpha_3 d_y + \epsilon_z \, ,$$

$$Y^* = \beta_0 + \beta_1 X + \epsilon_y \, ;$$

where

$$\text{cov } (\epsilon_z, \epsilon_y) = 0 \, ,$$

$$p(d_y = 1) = F(Y^*) \, , .$$

$$p(d_y = 0) = 1 - F(Y^*) \, .$$

Like model II, model IV can be reformulated to show explicitly the stochastic relationship between Y^* and d_y by introducing the continuous unobserved variable Y', which is related deterministically to d_y. Figure 6B shows this alternative formulation, and the equations of the model can be augmented to include the measurement equation $Y' = \lambda Y^* + \epsilon_{y'}$, where $d_y = 1$ if $Y' \geqslant M$, $d_y = 0$ if $Y' < M$. Model IV contains one more parameter (λ) than model III, reflecting the stochastic relationship between Y^* and d_y. As will be shown in the next section, model IV is not in general identifiable without additional information, although it can be identified in a wide range of circumstances.

The formal properties of models I–IV discussed thus far are summarized in the first two columns of table 1. Substantive differences among the models can be appreciated from two examples.

Consider a simple recursive model in which an individual's parents' socioeconomic status influences his grades of formal schooling achieved and both socioeconomic status and schooling affect earnings. Suppose, however, that the analyst does not measure schooling as a continuous variable but measures only whether or not individuals graduated from high school. This binary measure of schooling may capture the variation in schooling that affects earnings. Either the credential of a high school degree or the human capital derived from completion of all high school requirements may be the primary determinant of earnings (e.g., Jencks et al. 1979). Alternatively, however, the effect of grades of schooling achieved may be linear, with no bonus derived from high school completion. Given the data, these alternative hypotheses could be formulated using models I and III above. In figure 3, let Z be earnings, X be socioeconomic status, Y^* be an unmeasured variable interpreted as a continuous measure of schooling, and d_y be the observed dichotomy, graduate versus nongraduate. Under model I, schooling affects earnings as an unobserved continuous variable. Under the alternative hypothesis (model III, fig. 5), the variables denote the same concepts, but schooling affects earnings as an observed discrete variable.

A second example is the task of determining whether or not prior robbery victimization affects the likelihood of gun ownership when other

TABLE 1

Properties of Alternative Recursive Models with Discrete Intervening Variable (Y^*)

Model	Relationship between Discrete Outcome (d_y) and Unmeasured Continuous Variable (Y^*)	Causal Form of Intervening Variable	Identifiable from Simple Single-Equation Estimates	Identifiable from Reduced-Form Error Correlation	Identifiable with Instrumental Variables	Identifiable with Multiple Indicators of Y^*
I..............	Deterministic	Unobserved continuous variable (Y^*)	No	Yes	Yes	Yes
II..............	Stochastic	Unobserved continuous variable (Y^*)	No	No	Yes	Yes
III..............	Deterministic	Observed dummy variable (d_y)	Yes	Yes	Yes	Yes
IV..............	Stochastic	Observed dummy variable (d_y)	No	No	No	Yes

factors correlated with victimization, such as the crime rate in an individual's neighborhood, may also be important causes. If whether or not an individual owns a gun, whether or not he has recently been a victim, and other personal background factors are known, but neighborhood factors are unmeasured, then model II can be used to represent the alternative mechanisms through which victimization might affect ownership; to wit, either directly or as an indicator of the neighborhood climate that the individual experiences. In figure 4A and B, let X denote measured personal background, Z denote gun ownership, Y^* denote the neighborhood climate experienced by the individual, and d_y denote whether or not the individual has recently been a robbery victim. Under this model, gun ownership is affected by a latent continuous variable that is indicated by an observed discrete variable. In contrast, model IV (fig. 6A and B) represents a direct effect of observed victimization on ownership. In this example, gun ownership is a discrete variable. Strictly speaking, figures 4 and 6 should be modified to indicate that Z is an unobserved continuous variable related linearly to the predetermined variables and related nonlinearly to a discrete variable, say d_z. This modification, however, would leave unchanged the logical issues involved in considering the effects of Y^* and d_y. Models I and III might also be used to explore these arguments. Insofar as models II and IV are identifiable, however, they are most likely preferable. The latter models allow the relationship between neighborhood climate and personal experience to be stochastic. That is, some residents of dangerous neighborhoods have not been victims, and conversely, some residents of safe neighborhoods have been victims. Models I and III, in contrast, assume that whether or not an individual has been a victim exactly differentiates neighborhoods by their safety.

These examples raise the issue of testing rigorously among the alternative models. In practice, this requires that a more general model, one in which both the continuous and discrete effects of Y are estimated, be compared to the two more restrictive models (either I and III or II and IV). This procedure is illustrated in Section IV.

Of the four models considered in this section, model IV is closest to the recursive systems of logit equations for the causal analysis of discrete data discussed by Goodman (1972, 1973a, 1973b, 1979) and others. In these systems, when the discrete intervening variable is a dependent variable, it is formulated as the logit of the unobserved probability of an outcome on the discrete variable.[9] When the discrete variable is an in-

---

[9] The models discussed by Goodman (e.g., 1979) are equivalent to model IV if $\epsilon_y = 0$ and Y^* is viewed as a transformed unobservable probability p that has no substantive content beyond that contained in p itself. Model III can also be estimated as a log-linear or logit model, but here Y^* is viewed as an unobserved continuous variable with discrete indicator d_y. This is a different formulation from that of Goodman.

dependent variable, it is a dummy variable. Thus these methods for analysis of recursive systems are best applied to the substantive contexts for which model IV is appropriate.[10]

II. MODEL IDENTIFICATION

This section discusses the identification of recursive models for discrete data typified by models I–IV. First it examines the problem of identifying parameters in these models, focusing initially on single-equation models for discrete response variables and then on models I–IV in turn. The discussion of model identification assumes that equations with continuous response variables can be estimated by ordinary least squares (OLS), that equations with binary response variables can be estimated by logit or probit analysis, and that models with two or more equations can be estimated by extensions of these procedures. These methods of estimation are discussed in the Appendix.

Identification of parameters in structural equation models with discrete variables comprises two parts, the identification of parameters in a single equation with a continuous unobserved dependent variable indicated by an observed dichotomous variable, and the identification of parameters in multiequation models.

1. Single-Equation Models with Binary Dependent Variables

The preceding section discussed models in which discrete variables are indicators of unobserved continuous dependent variables, a formulation that was termed the "threshold model." As before, let d_y be a discrete response variable; X and W be observed independent variables; β_0, β_1, and β_2 be parameters; Y^* be an unobserved continuous variable; and ϵ_y be a stochastic disturbance with variance $\sigma_{\epsilon_y}^2$ assumed to be uncorrelated with X or W. Following the earlier discussion of the threshold model, $Y^* = \beta_0 + \beta_1 X + \beta_2 W + \epsilon_y$, where $d_y = 1$ if $Y^* \geq L$, $d_y = 0$ if $Y^* < L$, and L is the threshold parameter.

Without further specification, this model is underidentified because the scale of the unobserved variable Y^*, and thus of ϵ_y, is unknown. Under

---

[10] In practice, issues of identifiability and the treatment of the stochastic disturbances in these models are not addressed explicitly. Typically, it is assumed in the notation of model IV that $\lambda = 1$ and $\epsilon_y = 0$. Given the binomial trials interpretation that underpins such models, however, the parameters estimated under these systems of logit equations for the effect of X on Y are best interpreted as $\lambda\beta$, i.e., the product of the structural and measurement parameters. With sufficient information, however, the measurement and structural parts of model IV can be identified.

either a logit or probit specification, the estimated model is $y = b_0 + b_1 X + b_2 W + e_y$, where $y = Y^*/\sigma_{e_y}$, $b_0 = \beta_0/\sigma_{e_y}$, $b_2 = \beta_2/\sigma_{e_y}$, and $e_y = \epsilon_y/\sigma_{e_y}$. To identify the structural parameters β_0, β_1, β_2, and σ_{e_y}, an assumption about the variance of Y^* or ϵ_y is required. For example, if $\sigma_{e_y}^2 = 1$, then $\beta_0 = b_0$, $\beta_1 = b_1$, and $\beta_2 = b_2$. An alternative scaling assumption, employed throughout this article, is that var $(Y^*) = 1$. Under this assumption,

$$1 = \beta_1^2 \text{ var } (X) + \beta_2^2 \text{ var } (W) + 2\beta_1\beta_2 \text{ cov } (X,W) + \sigma_{e_y}^2$$

$$= \sigma_{e_y}^2 b_1^2 \text{ var } (X) + \sigma_{e_y}^2 b_2^2 \text{ var } (W) + 2\sigma_{e_y}^2 b_1 b_2 \text{ cov } (X,W) + \sigma_{e_y}^2 ,$$

and thus the parameters of interest can be obtained from the estimated parameters with the equations

$$\sigma_{e_y} = \sqrt{1/[b_1^2 \text{ var } (X) + b_2^2 \text{ var } (W) + 2b_1 b_2 \text{ cov } (X,W) + 1]} ;$$

$$\beta_0 = b_0\sigma_{e_y} ; \qquad \beta_1 = b_1\sigma_{e_y} ; \qquad \beta_2 = b_2\sigma_{e_y} .$$

In contrast to models with observed continuous dependent variables, therefore, a scaling assumption is required to identify model parameters.[11] With the scaling assumption used here, the slope parameters β_1 and β_2 measure the effects of a unit change in the independent variables X and W measured in standard deviations of the latent variable Y^*. Similar arguments are necessary to identify parameters in multiequation models discussed below.[12]

The binomial trials model presents identification problems in addition to that of scale identification. Consider again the threshold representation of the binomial trials model, that is, as in equations (8) and (9) above: $Y^* = \beta_0 + \beta_1 X + \epsilon_y$, $Y' = \lambda Y^* + \epsilon_{y'}$, where $d_y = 1$ if $Y' \geq M$, $d_y = 0$ if $Y' < M$. Because this model contains an additional source of variation, namely, $\epsilon_{y'}$, and an additional parameter λ, assumptions about the variances Y^* and Y' are insufficient to identify the model. Assume that both Y^* and Y' have variances of unity. Substitution of (8) into (9) yields $Y' = \lambda\beta_0 + \lambda\beta_1 X + \lambda\epsilon_y + \epsilon_{y'}$. Because this equation can be estimated

[11] This discussion has used the notation of the threshold model (eqq. [1] and [2]). The same issue of scale identification arises in the binomial trials model (eqq. [3]–[9]), although the additional parameter λ in the latter model raises additional identification problems (see below).

[12] A further requirement for the identification of single-equation models is that the dichotomous dependent variable does not classify perfectly any of the independent variables or linear combinations of the independent variables; i.e., there can be no value of an independent variable above which all values of d_y are 1 and below which all values are 0 (Heckman 1978).

directly through logit or probit analysis, the quantities $\lambda\beta_0$ and $\lambda\beta_1$ are identified, but the separate structural and measurement parameters β_0, β_1, and λ are not.

If one assumes, however, that Y^* is an exact linear function of X, that is, that $\epsilon_y = 0$ for all observations, it is possible to identify both the structural and measurement parameters. Substitution of (8) into (9) assuming $\epsilon_y = 0$ yields $Y' = \lambda\beta_0 + \lambda\beta_1 X + \epsilon_{y'}$, which can be estimated by probit or logit analysis. As in the case of the threshold model, however, the estimated equation is $Y' = c_0 + c_1 X + e_{y'}$, where $y' = Y'/\sigma_{\epsilon_{y'}}$, $c_0 = \lambda\beta_0/\sigma_{\epsilon_{y'}}$, $c_1 = \lambda\beta_1/\sigma_{\epsilon_{y'}}$, $e_{y'} = \epsilon_{y'}/\sigma_{\epsilon_{y'}}$, and $\sigma_{\epsilon_{y'}}^2 = \text{var}(\epsilon_{y'})$. If $\text{var}(Y') = 1$, then $\sigma_{\epsilon_{y'}}^2 = 1/[1 + c_1^2 \text{ var}(X)]$ and

$$\lambda = 1 - \sqrt{1/[1 + c_1^2 \text{ var}(X)]} \, .$$

The structural parameters can be obtained as follows: $\beta_0 = c_0\sigma_{\epsilon_{y'}}/\lambda$; $\beta_1 = c_1\sigma_{\epsilon_{y'}}/\lambda$. Alternative strategies are discussed below. One means of identifying both the structural and measurement parts of the binomial trials model is to employ multiple indicators of Y^* (Muthén 1979). This approach is illustrated for model IV below. Another approach is to use additional information in the form of an instrumental variable. This is illustrated for model III.

2. Model I

Consider again the structural equations of model I:

$$Z = \alpha_0 + \alpha_1 X + \alpha_2 Y^* + \epsilon_z \, , \tag{10}$$

$$Y^* = \beta_0 + \beta_1 X + \epsilon_y \, ; \tag{11}$$

where $d_y = 1$ if $Y^* \geq L$, $d_y = 0$ if $Y^* < L$, Y^* is a continuous unobserved variable, $\text{cov}(\epsilon_z, \epsilon_y) = 0$, $\text{var}(Y^*) = 1$, $\text{var}(\epsilon_z) = \sigma_{\epsilon_z}^2$, and $\text{var}(\epsilon_y) = \sigma_{\epsilon_y}^2$. Equation (11) can be estimated directly by probit or logit analysis, given the scale identification assumption on Y^*. Thus β_0, β_1, and $\sigma_{\epsilon_y}^2$ are identifiable. Equation (10), however, creates additional problems because Y^* is unobserved and thus standard estimation procedures for either discrete or continuous measures of Z cannot be applied directly. Equation (10) is nonetheless identified, as can be seen by substituting (11) into (10) to yield the reduced-form equation

$$Z = \gamma_0 + \gamma_1 X + \eta_z \, , \tag{12}$$

where

$$\gamma_0 = \alpha_0 + \alpha_2\beta_0 \, , \tag{13}$$

$$\gamma_1 = \alpha_1 + \alpha_2\beta_1 , \tag{14}$$

$$\eta_s = \epsilon_s + \alpha_2\epsilon_y , \tag{15}$$

$$\text{var}(\eta_s) = \sigma_{\eta_s}^2 = \sigma_{\epsilon_s}^2 + \alpha_2^2\sigma_{\epsilon_y}^2 ,$$

and

$$\text{cov}(\eta_s,\epsilon_y) = \alpha_2\sigma_{\epsilon_y}^2 . \tag{16}$$

If Z is continuous, (12) can be estimated by OLS; if Z is discrete, (12) can be estimated by probit or logit analysis. As is shown in the Appendix, if an assumption is made about the multivariate distribution of ϵ_s and ϵ_y, and thus about η_s, it is possible to estimate $\text{cov}(\eta_s,\epsilon_y)$. Thus (13), (14), and (16) provide three equations in the three unknown parameters α_0, α_1, and α_2. The parameters of (10), therefore can be identified from the equations:

$$\alpha_0 = \gamma_0 - \beta_0 \left[\frac{\text{cov}(\eta_s,\epsilon_y)}{\sigma_{\epsilon_y}^2} \right] ,$$

$$\alpha_1 = \gamma_1 - \beta_1 \left[\frac{\text{cov}(\eta_s,\epsilon_y)}{\sigma_{\epsilon_y}^2} \right] ,$$

$$\alpha_2 = \frac{\text{cov}(\eta_s,\epsilon_y)}{\sigma_{\epsilon_y}^2} .$$

Thus, model I is identifiable from equations predicting Y^* and from the parameters of the reduced-form equation for Z.

This identification procedure requires that an estimate of the disturbance covariance, $\text{cov}(\eta_s,\epsilon_y)$, for (11) and (12) be obtained. Although this is always possible, it requires that one assume a particular multivariate distribution for ϵ_s and ϵ_y, and it is computationally burdensome.[13] In the absence of this estimate, model I cannot be identified. However, if there is additional information in the form of an instrumental variable for Y^*— that is, a variable, say W, that affects Y^* but does not affect Z directly—

[13] We are unaware of any research on the robustness under alternative distributional assumptions of the models considered here. As is well-known, maximum likelihood estimates are inconsistent if they rest on incorrect distributional assumptions, and it is usually preferable to adopt methods that are robust. Although it is applied to models quite different from those considered here, recent work by Heckman and Singer (1981) and by Goodman (1981) explores the implications of varying distributional assumptions and may lead to more robust methods.

then it is possible to identify the model without cov (η_z, ϵ_y). To see this, modify (11) as follows:

$$Y^* = \beta_0 + \beta_1 X + \beta_2 W + \epsilon_y .$$ (11*)

Then the reduced-form equation for Z (eq. [12]) becomes

$$Z = \gamma_0 + \gamma_1 X + \gamma_2 W + \eta_z ,$$ (12*)

where

$$\gamma_2 = \alpha_2 \beta_2 .$$ (17)

Equations (13), (14), and (17) contain the three parameters α_0, α_1, and α_2 and can be solved for them as follows: $\alpha_0 = \gamma_0 - (\beta_0 \gamma_2 / \beta_2)$; $\alpha_1 = \gamma_1 - (\beta_1 \gamma_2 / \beta_2)$; $\alpha_2 = \gamma_2 / \beta_2$.

3. Model II

Model II, in which the observed discrete variable d_y is related stochastically to an unmeasured continuous variable, contains one more parameter than model I. As a result, without additional information it is not identified. To see this, consider the model, which is diagramed in figure 4B:

$$Z = \alpha_0 + \alpha_1 X + \alpha_2 Y^* + \epsilon_z ,$$ (18)

$$Y^* = \beta_0 + \beta_1 X + \epsilon_y ,$$ (19)

$$Y' = \lambda Y^* + \epsilon_{y'} ;$$ (20)

where $d_y = 1$ if $Y' \geq M$, $d_y = 0$ if $Y' < M$, var $(Y') = $ var $(Y^*) = 1$, cov $(\epsilon_z, \epsilon_y) = $ cov $(\epsilon_y, \epsilon_{y'}) = $ cov $(\epsilon_z, \epsilon_{y'}) = 0$, var $(\epsilon_z) = \sigma_{\epsilon_z}^2$, var $(\epsilon_y) = \sigma_{\epsilon_y}^2$, var $(\epsilon_{y'}) = \sigma_{\epsilon_{y'}}^2$, and all other notation is as described above. Equations (18), (19), and (20) can be expressed in reduced form for Z and Y' as follows:

$$Z = \gamma_0 + \gamma_1 X + \eta_z ,$$ (21)

$$Y' = \theta_0 + \theta_1 X + \eta_{y'} ;$$ (22)

where

$$\gamma_0 = \alpha_0 + \alpha_2 \beta_0 ,$$ (23)

$$\gamma_1 = \alpha_1 + \alpha_2\beta_1 , \tag{24}$$

$$\theta_0 = \lambda\beta_0 , \tag{25}$$

$$\theta_1 = \lambda\beta_1 , \tag{26}$$

$$\text{cov} (\eta_z, \eta_{y'}) = \text{cov} (\epsilon_{y'} + \lambda\epsilon_y, \epsilon_z + \alpha_2\epsilon_y)$$
$$= \alpha_2\lambda\sigma_{\epsilon_y}^2 , \tag{27}$$

$$\text{var} (\eta_z) = \sigma_{\epsilon_z}^2 + \alpha_2^2\sigma_{\epsilon_y}^2 , \tag{28}$$

$$\text{var} (\eta_{y'}) = \sigma_{\epsilon_{y'}}^2 + \lambda^2\sigma_{\epsilon_y}^2 . \tag{29}$$

The reduced-form equations (21) and (22) can be estimated by OLS (for continuous Z) and either logit or probit analysis, respectively, providing estimates of γ_0, γ_1, θ_0, θ_1, cov $(\eta_z, \eta_{y'})$, var (η_z), and var $(\eta_{y'})$. These estimates, however, yield seven equations, (23)–(29), in eight unknown structural parameters: α_0, α_1, α_2, β_0, β_1, λ, σ_{ϵ_z}, and σ_{ϵ_y}. None of these parameters is identifiable. Note, however, that if $\lambda = 1$, model II reduces to model I and all parameters can be estimated as for the latter model.

Model II can, nonetheless, be identified in its own right with additional information. In particular, if there exists another exogenous variable, say W, that affects Y^* but not Z, model II can be identified. Alternatively, if there are two observed dichotomous variables that index Y^*, say d_{y_1} and d_{y_2}, instead of just one, then, too, the model can be identified. The following shows how an instrumental variable W can be used to identify model II. The use of multiple discrete indicators is outlined in the discussion for model IV below.

Modify model II by replacing (19) with

$$Y^* = \beta_0 + \beta_1 X + \beta_2 W + \epsilon_y \tag{19*}$$

but allowing (18) and (20) to remain as above. Then the reduced forms become

$$Z = \gamma_0 + \gamma_1 X + \gamma_2 W + \eta_z , \tag{21*}$$

$$Y' = \theta_0 + \theta_1 X + \theta_2 W + \eta_{y'} , \tag{22*}$$

where equations (23)–(29) hold as before and, in addition,

$$\theta_2 = \lambda\beta_2 , \tag{30}$$

$$\gamma_2 = \alpha_2\beta_2 . \tag{31}$$

Equations (23)–(31) now provide nine independent equations in the nine unknown structural parameters. These equations, along with the scaling assumptions on Y^* and Y', yield the following solutions for the structural parameters:

$$\lambda = \sqrt{\left[1 - \text{var } (\eta_{y'}) + \frac{\theta_2}{\gamma_2} \text{cov } (\eta_z, \eta_{y'})\right]} = k \; ;$$

$$\beta_0 = \frac{\theta_0}{k} \; ; \quad \beta_1 = \frac{\theta_1}{k} \; ; \quad \beta_2 = \frac{\theta_2}{k} \; ;$$

$$\alpha_0 = \gamma_0 - \frac{\gamma_2 \theta_0}{\theta_2} \; ; \quad \alpha_1 = \gamma_1 - \frac{\gamma_2 \theta_1}{\theta_2} \; ;$$

$$\alpha_2 = \frac{\gamma_2}{\theta_2} k \; ; \quad \sigma_{\epsilon_y} = \sqrt{\left[\frac{\theta_2}{\gamma_2} \text{cov } (\eta_z, \eta_{y'})\right]} / k \; ;$$

$$\sigma_{\epsilon_z} = \sqrt{\text{var } (\eta_z) - \frac{\gamma_2}{\theta_2} \text{cov } (\eta_z, \eta_{y'})} \; .$$

Thus the introduction of an instrumental variable W that affects Y^* but not Z allows the identification of all structural parameters. The single restriction that W does not affect Z compensates for the single additional parameter λ included in model II but not model I.

4. Model III

Recall that in model III the observed discrete response d_y is related deterministically to a latent continuous variable Y^* and has a direct impact on the endogenous variable Z (see fig. 5). That is, $Z = \alpha_0 + \alpha_1 X + \alpha_3 d_y + \epsilon_z$ and $Y^* = \beta_0 + \beta_1 X + \epsilon_y$, where var $(Y^*) = 1$, cov $(\epsilon_z, \epsilon_y) = 0$, and all notation is as defined above. Since all the independent variables are observed, model III can be estimated by the usual single-equation methods; that is, OLS for continuous Z or logit or probit analysis for discrete Z and logit or probit analysis for Y^*. Subject to the scale restrictions on Y^*, the model is identified directly and presents no special problems.

5. Model IV

Finally, consider model IV, which differs from model III only in that the relationship between the observed discrete variable and the latent con-

tinuous variable to which it corresponds is stochastic. This model, diagramed in figure 6B, is written as above:

$$Z = \alpha_0 + \alpha_1 X + \alpha_3 d_y + \epsilon_z , \tag{32}$$

$$Y^* = \beta_0 + \beta_1 X + \epsilon_y , \tag{33}$$

$$Y' = \lambda Y^* + \epsilon_{y'} ; \tag{34}$$

where $d_y = 1$ if $Y' \geq M$, $d_y = 0$ if $Y' < M$, cov $(\epsilon_z, \epsilon_y) =$ cov $(\epsilon_y, \epsilon_{y'}) =$ cov $(\epsilon_z, \epsilon_{y'}) = 0$, var $(Y') =$ var $(Y^*) = 1$, and all other notation is as described above. Equation (32) can be estimated directly by OLS if Z is continuous or by logit or probit analysis if Z is discrete. Equations (33) and (34), however, are not identifiable. To see this, consider the reduced-form equation for Y', $Y' = \theta_0 + \theta_1 X + \eta_{y'}$, where $\theta_0 = \lambda\beta_0$, $\theta_1 = \lambda\beta_1$, and var $(\eta_{y'}) = \sigma_{\epsilon_{y'}}^2 + \lambda^2\sigma_{\epsilon_y}^2$. Since cov $(\eta_{y'}, \epsilon_z) = 0$, there are only three equations from which to estimate the four parameters λ, β_0, β_1, and $\sigma_{\epsilon_y}^2$. A key feature of this model is that, although the measurement parameter λ cannot be separated from the structural parameter β_1, the effect of X on Z via Y^* can nonetheless be calculated from the reduced-form parameter θ_1. This will be shown in the next section.

Unlike model II, which is also underidentified without additional information, model IV cannot be identified by resort to instrumental variables, that is, variables affecting Y^* but not Z. An alternative method that does permit identification, however, is to obtain additional categorical indicators of Y^*. Muthén (1979) shows that repeated observations on Y^* can be used to identify models similar to model IV. Consider an extension of model IV which retains (33) but replaces (32) and (34) with the three equations:

$$Z = \alpha_0 + \alpha_1 X + \alpha_3 d_{y_1} + \alpha_4 d_{y_2} + \epsilon_z , \tag{32*}$$

$$Y'_1 = \lambda_1 Y^* + \epsilon_{y'_1} , \tag{35}$$

$$Y'_2 = \lambda_2 Y^* + \epsilon_{y'_2} ; \tag{36}$$

where $d_{y_1} = 1$ if $Y'_1 \geq M_1$, $d_{y_1} = 0$ if $Y'_1 < M_1$, $d_{y_2} = 1$ if $Y'_2 \geq M_2$; $d_{y_2} = 0$ if $Y'_2 < M_2$; var $(Y'_1) =$ var $(Y'_2) = 1$; ϵ_z, $\epsilon_{y'_1}$, and $\epsilon_{y'_2}$ are uncorrelated; M_1 and M_2 are threshold parameters; and

$$\text{var } (\epsilon_{y'_1}) = \sigma_{\epsilon_{y'_1}}^2 \quad \text{and} \quad \text{var } (\epsilon_{y'_2}) = \sigma_{\epsilon_{y'_2}}^2 .$$

This model, which is diagramed in figure 7, contains two observed dichotomous indicators, d_{y_1} and d_{y_2}, of the underlying construct Y^*. As in

the single-indicator case, Y'_1 and Y'_2 are related linearly but stochastically to Y^* but related deterministically to d_{y_1} and d_{y_2}, respectively. This model might be applied to an extension of the robbery victimization–gun ownership example discussed above. Let X denote personal background characteristics, Y^* denote neighborhood climate, d_{y_1} denote whether the individual has recently been robbed, d_{y_2} denote whether or not the individual knows someone else who has recently been robbed, and Z denote gun ownership. Model IV then models the effects of personal experiences with robbery on ownership.

This model can be identified as follows. Again the equation for Z can be estimated directly using a method appropriate to the way that Z is measured. To obtain the remaining parameters, solve (35) and (36) to obtain the reduced-form equations for Y'_1 and Y'_2: $Y'_1 = \theta_0 + \theta_1 X + \eta_{y'_1}$ and $Y'_2 = \psi_0 + \psi_1 X + \eta_{y'_2}$, where

$$\theta_0 = \beta_0\lambda_1 ; \quad \theta_1 = \beta_1\lambda_1 ; \quad \psi_0 = \beta_0\lambda_2 ; \quad \psi_1 = \beta_1\lambda_2 ;$$

$$\mathrm{cov}\,(\eta_{y'_1},\eta_{y'_2}) = \lambda_1\lambda_2\sigma_\epsilon^2 ; \quad \mathrm{var}\,(\eta_{y'_1}) = \sigma_{\epsilon_{y'_1}}^2 + \lambda_1^2\sigma_\epsilon^2 ;$$

$$\mathrm{var}\,(\eta_{y'_2}) = \sigma_{\epsilon_{y'_2}}^2 + \lambda_2^2\sigma_\epsilon^2 .$$

These equations can be solved for the parameters of interest; that is,
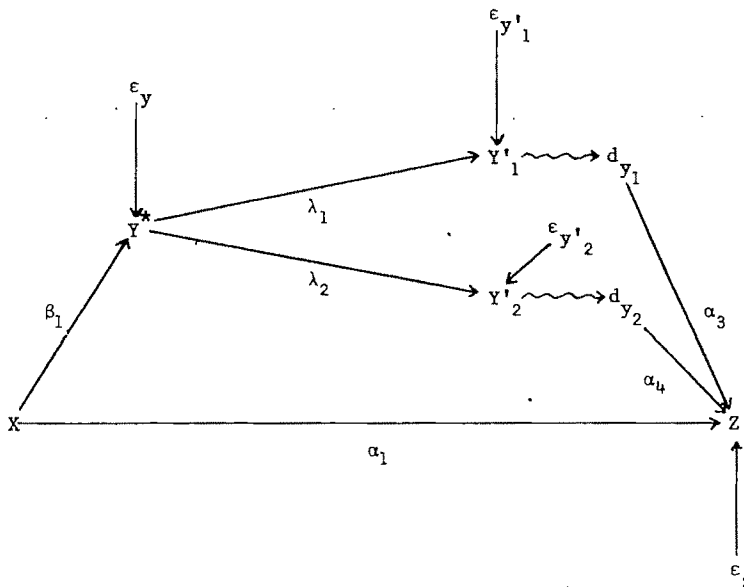
Fig. 7.—Threshold representation of model IV with multiple indicators

$$\lambda_1 = \sqrt{1 - \text{var}(\eta_{y'_1}) + \frac{\theta_1}{\psi_1}\,\text{cov}(\eta_{y'_1}, \eta_{y'_2})} = k_1\,,$$

$$\lambda_2 = \sqrt{1 - \text{var}(\eta_{y'_2}) + \frac{\psi_1}{\theta_1}\,\text{cov}(\eta_{y'_1}, \eta_{y'_2})} = k_2\,,$$

$$\beta_0 = \frac{\theta_0}{k_1}\,; \quad \beta_1 = \frac{\theta_1}{k_1}\,; \quad \sigma_\epsilon{}^2 = \frac{\text{cov}(\eta_{y'_1}, \eta_{y'_2})}{k_1 k_2}\,,$$

$$\sigma_{\epsilon_{y'_1}}{}^2 = 1 - k_1{}^2\,; \quad \sigma_{\epsilon_{y'_2}}{}^2 = 1 - k_2{}^2\,.$$

Thus multiple indicators of Y^* permit model IV to be identified.

The foregoing discussion has shown that the four basic recursive models vary considerably in their ease of identification. The alternative avenues to identification for these models are summarized in table 1.

Models I, II, III, and IV, of course, are not the only possible recursive models. An important type of model is one that combines the features of models I and III or those of models II and IV. That is, the effects of a discrete variable result from both the discrete variable, per se, and a latent continuous variable of which the discrete variable is an indicator. Indeed, such models are necessary to assess the relative importance of the features of models I and III or those of II and IV. It can be shown that a model that combines the features of models I and III can be identified under the same conditions as model I. That is, simple single-equation estimates will not suffice to identify the model, but using an estimate of the reduced-form error covariance (cov $[\eta_x, \epsilon_y]$), or an instrumental variable that either affects Y^* but not Z or has multiple indicators of Y^*, allows the model to be identified. Similarly, it can be shown that a model that combines the features of models II and IV can be identified under the same conditions as model IV. That is, multiple indicators of Y^* identify the model. These procedures are illustrated in Section IV below.

III. PATH ANALYSIS IN RECURSIVE MODELS WITH DISCRETE ENDOGENOUS VARIABLES

This section shows that the same principles of path analysis that have been applied to models with continuous dependent variables (Alwin and Hauser 1975; Duncan 1966) can be applied to the models considered in this article. That is, the total effect of a predetermined variable, say X, on an endogenous variable, say Z, can be apportioned into a component due to the indirect effect of X on Z through an intervening variable, say

Y, and a direct effect of X on Z net of Y. The "calculus of paths" is slightly more complex when Z or Y is discrete than when both are continuous because not all the relationships in the model are linear. Stolzenberg (1979), however, shows that such calculations are feasible for nonlinear models in general, and this section applies his methods to nonlinearities arising in models with discrete variables.

1. Continuous Endogenous Variables

Consider again the two-equation model for the ith observation: $Z = \alpha_0 + \alpha_1 X + \alpha_2 Y + \epsilon_z$, and $Y = \beta_0 + \beta_1 X + \epsilon_y$, where cov $(\epsilon_z, \epsilon_y) = 0$. If both Z and Y are continuous, the direct and indirect effects of X and Z are given by

$$
\begin{array}{cccc}
b_{zx} & = & \alpha_1 & + & \alpha_2 \beta_1 \, , \\
\text{(total)} & & \text{(direct)} & & \text{(indirect)}
\end{array}
\qquad (37)
$$

where b_{zx} is the zero-order (and reduced-form) regression of Z on X. This relationship holds whether Z and Y are observed continuous variables or latent variables with dichotomous indicators, say d_z and d_y. Thus (37) is an appropriate decomposition of the direct and indirect effects of X on Z for models I and II (see figs. 3 and 4), in which the effect of Y is hypothesized to be that of a latent continuous variable.

2. Continuous Intervening Variable; Discrete Z

The decomposition (37) is measured in units of the continuous variable Z, whether observed or unobserved. In any of models I through IV, however, when Z is a latent variable corresponding to a discrete variable d_z, the analyst may wish to calculate the effects of X on Z in terms of the discrete variable d_z rather than the continuous variable itself. It may, for example, be more meaningful to determine the effects of X on the probability that $d_z = 1$ than on the logit or probit transform of that probability. In this case, and in models III and IV, where the effect of Y is hypothesized to be that of an observed discrete variable d_y, (37) must be modified to take into account the nonlinear relationships between the discrete and continuous variables.

By elementary calculus, a dependent variable Z may be decomposed by its total differential $dZ = (\partial Z/\partial X)dX + (\partial Z/\partial Y)dY$, which implies the general formula

$$
\frac{dZ}{dX} = \frac{\partial Z}{\partial X} + \frac{\partial Z}{\partial Y}\frac{dY}{dX} \, ,
\qquad (38)
$$

83

of which (37) is a·special case (Stolzenberg 1979). Now let Z be a latent variable indicated by an observed dichotomous variable d_z and let Y be a continuous variable. This is again model I or II. It is possible to calculate the expected effects of X on d_z, that is, the effects of X on the expected values of d_z (given X).[14] The expected values of d_z are given by the equation $p(d_z = 1) = F(\alpha_0 + \alpha_1 X + \alpha_2 Y)$, where, as before, F is a general function that is typically the logistic or cumulative normal. Then the total effect of X on $p(d_z = 1)$ is, using (38), decomposed as

$$\frac{dp(d_z = 1)}{dX} = \alpha_1 f(\alpha_0 + \alpha_1 X + \alpha_2 Y) + \alpha_2 [f(\alpha_0 + \alpha_1 X + \alpha_2 Y)] \frac{dY}{dX}$$

$$= \alpha_1 f(\alpha_0 + \alpha_1 X + \alpha_2 Y) + \alpha_2 [f(\alpha_0 + \alpha_1 X + \alpha_2 Y)] \beta_1 , \quad (39)$$

$$\text{(total)} \qquad\qquad \text{(direct)} \qquad\qquad \text{(indirect)}$$

where f is the derivative of F, that is, the probability density function corresponding to F (Hanushek and Jackson 1977). If F is the logistic function,

$$\frac{dp(d_z = 1)}{dX} = \frac{\alpha_1 \exp (\alpha_0 + \alpha_1 X + \alpha_2 Y)}{[1 + \exp (\alpha_0 + \alpha_1 X + \alpha_2 Y)]^2} + \frac{\alpha_2 \beta_1 \exp (\alpha_0 + \alpha_1 X + \alpha_2 Y)}{[1 + \exp (\alpha_0 + \alpha_1 X + \alpha_2 Y)]^2} \quad (40)$$

$$= \alpha_1 p_z (1 - p_z) + \alpha_2 \beta_1 p_z (1 - p_z) ,$$

where $p_z = p(d_z = 1)$ is the probability that $d_z = 1$. If F is the cumulative normal distribution function,

$$\frac{dp(d_z = 1)}{dX} = \frac{\alpha_1 \exp (-t^2/2)}{\sqrt{2\pi}} + \frac{\alpha_2 \beta_1 \exp (-t^2/2)}{\sqrt{2\pi}} , \quad (41)$$

where t is the standardized normal variable corresponding to $p(d_z = 1)$ under the probit transformation. The key feature of these decompositions

[14] The motivation for examining "expected effects" of independent variables on dichotomous dependent variables differs somewhat under the binomial trials and threshold views of discrete variables. In the binomial trials model, where each observation has an unobserved probability, say p, that it will take the value 1 on d_y, the expected effect can be interpreted as the expected change in p from observation to observation for a change in the value of X. In the threshold model, in contrast, changes in X change d_y only when Y^* is at its threshold L; i.e., $d(d_y)/dY^* = 0$ if $Y^* \neq L$, $d(d_y)/dY^* = \pm \infty$ if $Y^* = L$. In this case the expected effect can be interpreted as the effect an observer might expect for a group of observations with a common value on X; i.e., the effect is the average for the population of observations with the same values on X. If the observations refer to individuals, then, in the binomial trials model, the estimated effects are those that can be expected for both the individual and the investigator. In the threshold model, the estimated effects are expected only by the investigator who examines a group of individuals with a common value on X.

84

is that they depend on the observations being considered; that is, the effects of X on d_z depend on the level of X. This is, of course, exactly what the nonlinear hypothesis of the probit or logit model requires. In practice, therefore, one should calculate the effects of X for a range of values of X or p to chart the nonlinear relationship. Among other values, perhaps the most useful summary calculation is to choose the sample proportion of cases for which $d_z = 1$. This proportion can be substituted for p_z in (40) or its normal probability transform t can be substituted in (41). Note, however, that the *relative* importance of the direct and indirect effects is invariant under the choice of X or of p because $f(\alpha_0 + \alpha_1 X + \alpha_2 Y)$ is simply a constant of proportionality in (39). Moreover, for models I and II, the relative importance will be the same whether the effects of X are decomposed in the scale of the latent continuous variable Z as in (37) or in the scale of $p(d_z = 1)$ as in (39).

3. Discrete Intervening Variable

The preceding discussion assumed that the intervening variable Y affected Z as a continuous variable, either latent or observed. Now consider the decomposition for models III and IV (figs. 5 and 6) where the discrete variable d_y affects Z directly. The model is $Z = \alpha_0 + \alpha_1 X + \alpha_3 d_y + \epsilon_z$, and $Y = \beta_0 + \beta_1 X + \epsilon_y$, where cov $(\epsilon_z, \epsilon_y) = 0$.

Model III.—For decomposition it is necessary to take account of the nonlinear relationship between d_y and X. The expected value of d_y is $p(d_y = 1) = F(\beta_0 + \beta_1 X)$, where F is typically the logistic or cumulative normal function, and thus the expected effect of X on d_y is given by $dp(d_y = 1)/dX = \beta_1 f(\beta_0 + \beta_1 X)$, where f is the probability density function, either extreme value or normal, corresponding to F (Hanushek and Jackson 1977). Now if Z is continuous, the effect of X on Z[15] is

$$
\begin{aligned}
\frac{dZ}{dX} &= \alpha_1 + \frac{\alpha_3 dp(d_y = 1)}{dX} \\
&= \alpha_1 + \alpha_3 \beta_1 f(\beta_0 + \beta_1 X) .
\end{aligned}
\tag{42}
$$

(total)   (direct)   (indirect)

If F is logistic, $dZ/dX = \alpha_1 + \alpha_3 \beta_1 p_y (1 - p_y)$, where $p_y = p(d_y = 1)$ and $p_y = \exp(\beta_0 + \beta_1 X)/[1 + \exp(\beta_0 + \beta_1 X)]$. If F is the cumulative normal function, $dZ/dX = \alpha_1 + \alpha_3 \beta_1 [\exp(-t^2/2)/\sqrt{2\pi}]$. If Z is a discrete

---

[15] More precisely, this formula gives the *expected* effect of X on Z, i.e., $dE(Z)/dX$, inasmuch as the indirect effect of X on Z includes the effect of X on $p(d_y = 1)$. For expository convenience, this mathematical nuance is ignored in the formulas presented in this section.

variable and the effects of X on the expected values of d_z are required, the decomposition is

$$\frac{dp(d_z = 1)}{dX} = \underbrace{\alpha_1 f(\alpha_0 + \alpha_1 X + \alpha_3 d_y)}_{\text{(direct)}}$$

(total)   (direct)   (43)

$$+ \underbrace{\alpha_3 f(\alpha_0 + \alpha_1 X + \alpha_3 d_y)\beta_1 f(\beta_0 + \beta_1 X)}_{\text{(indirect)}}.$$

(indirect)

As before, f is the extreme value density function under a logit model or the standard normal density function under the probit model. In the former case, for example, the decomposition is $dp(d_z = 1)/dX = \alpha_1 p_z(1 - p_z) + \alpha_3 p_z(1 - p_z)\beta_1 p_y(1 - p_y)$, where specific values of $p_z = p(d_z = 1)$ and $p_y = p(d_y = 1)$, say the sample proportions for which $d_z = 1$ and for which $d_y = 1$, must be chosen.

Model IV.—In addition to the structural equations given above, model IV includes the measurement equation $Y' = \lambda Y + \epsilon_{y'}$, indicating that the link between Y and d_z is stochastic as well as nonlinear. Under this model, $dp(d_y = 1)/dX = [dp(d_y = 1)/dY]\,dY/dX = \beta_1\lambda f(\beta_0 + \beta_1 X)$. If Z is continuous, the decomposition of direct and indirect effects of X on Z is

$$\frac{dZ}{dX} = \alpha_1 + \alpha_3\beta_1\lambda f(\beta_0 + \beta_1 X).$$

(total)   (direct)   (indirect)

If the effects of X on d_z are required, the decomposition is

$$\frac{dp(d_z = 1)}{dX} = \alpha_1 f(\alpha_0 + \alpha_1 X + \alpha_3 d_y)$$

(total)   (direct)

$$+ \alpha_3[f(\alpha_0 + \alpha_1 X + \alpha_3 d_y)]\beta_1\lambda[f(\beta_0 + \beta_1 X)].$$

(indirect)

Here and in (42) and (43) the relative sizes of the direct and indirect components are unchanged whether one calculates the effects on Z or on $p(d_z = 1)$. Note also that for model IV it is unnecessary to identify the measurement parameter λ to calculate the path decomposition. It suffices to be able to identify $\beta_1\lambda$ since only this product is used in the calculations. Thus the decomposition for model IV is feasible under broader circumstances than those in which the model itself can be identified.

IV. EMPIRICAL EXAMPLES

This section illustrates the models and methods presented above with two numerical examples. The first applies models I and III to the effects of social background and parental factors on college plans of high school seniors. The second applies models II and IV to the effects of state Fair-Employment-Practices Legislation and other social factors on the relative earnings of black and white workers.

1. Social Background, Parental Encouragement, and College Plans

Using a sample of high school seniors, Sewell and Shah (1968) and Fienberg (1980) examine the effects of social background factors, including sex, socioeconomic status, and intelligence, on the extent to which their parents encourage them to attend college, and the effects of both social background and parental encouragement on whether the seniors themselves plan to attend college. The researchers apply a simple recursive model with two endogenous variables: parental encouragement and college plans. Encouragement is interpreted as intervening causally between social background and plans. Both endogenous variables are measured discretely. Whether or not the senior plans to attend college is coded yes or no. Parental encouragement to attend college is coded as high or low. Previous analyses of these data have examined the effect on college plans of a dummy variable for parental encouragement, with the social background variables controlled, and the effects of the background variables on parental encouragement. This two-equation model, therefore, is an instance of either model III or model IV discussed above. Parental encouragement, however, may be better modeled as resulting from variation on an unmeasured continuous variable indicating that degree of encouragement is a continuum from none to a great deal, than as a simple binary difference between those coded as experiencing "high" or "low" encouragement. In this case, the recursive model of choice is either model I or model II discussed above since these models allow the effects of discrete intervening variables to result from continuous latent variables. These models do not necessarily fit the data equally well. Nor do they have the same implications for the degree to which parental encouragement mediates the effect of social background on college plans.

Data.—The data are a cross-classification of the characteristics of 10,318 Wisconsin high school seniors in 1957. The characteristics and their measurement are as follows: socioeconomic status (S) (high, upper-middle, lower-middle, low), intelligence (Q) as measured by the Hammon-Nelson Test of Mental Ability (high, upper-middle, lower-middle, low), sex (X) (female, male), parental encouragement (P) (high, low), and college plans

(C) (yes, no). The data are reported by Sewell and Shah (1968) and by Fienberg (1980, p. 130).[16] In the models estimated here, both IQ and socioeconomic status are coded as continuous variables scaled to have means of zero and variances of unity.[17]

Models.—As noted above, it is of interest to contrast alternative roles of parental encouragement in recursive models of background effects on college plans. To carry this out, models I and III and a model that combines the effects of models I and III are estimated. Models I and III are nested within the latter model, and thus their relative fits to the data can be derived from the three models. Models I and III rather than models II and IV are considered here because the former are identifiable without additional assumptions, whereas the latter are not. In modeling parental encouragement, moreover, models I and III are substantively plausible. That is, the continuous underlying variable for parental encouragement may be reasonably assumed to be classified perfectly by the observed dichotomous variable. That is, there are no individuals with high scores on the unmeasured encouragement measure who are coded "low" on the observed variable or vice versa. Measurement error is, of course, always possible, but there is no other substantive reason for modeling the link between the continuous and discrete variables as stochastic.

For models I, III, and their combination, the effects of social background can be modeled as $P^* = \beta_0 + \beta_1 X + \beta_2 Q + \beta_3 S + \epsilon_p$, where ϵ_p is a normally distributed stochastic disturbance, P^* is an unmeasured continuous variable, and $P =$ "high" if $P^* \geq L$, $P =$ "low" if $P^* < L$, where L is an unobserved threshold. For model I the effects of background and parental encouragement can be written $C^* = \alpha_0 + \alpha_1 X + \alpha_2 Q + \alpha_3 S + \alpha_4 P^* + \epsilon_c$, where ϵ_c is a normally distributed disturbance, C^* is an unmeasured continuous variable, and $C =$ "yes" if $C^* \geq N$, $C =$ "no" if $C^* < N$, where N is an unobserved threshold. For model III the effect of P on C is discrete, and the model can be written $C^* = \alpha_0 + \alpha_1 X + \alpha_2 Q + \alpha_3 S + \alpha_5 P + \epsilon_c$. Finally, the combined model is $C^* = \alpha_0 +$

---

[16] The data are also available from us on request.

[17] Fienberg's (1980) reported results as well as additional analyses of the data show that some of the effects of socioeconomic status and intelligence in the model are nonlinear. Fienberg also reports significant three-way interactions among pairs of the background variables and the endogenous variables. The models reported here assume both linearity of the background effects and additivity of all effects in the model. These simplifications clarify the exposition and yield substantively plausible results. The models presented here do not fit the data as well as the more complex models estimated by Fienberg, and the differences in fit between models I and III reported below *may* result in part from failure to fit higher-order interactions to the data. It is nonetheless straightforward to incorporate nonlinearities and higher-order interactions into the models presented here. However, in practice, models with higher-order interactions and models with latent continuous variables may be empirically indistinguishable without prior conceptions of which specification is more plausible.

$\alpha_1 X + \alpha_2 Q + \alpha_3 S + \alpha_4 P^* + \alpha_5 P + \epsilon_c$. For all models, ϵ_c and ϵ_p are assumed to be uncorrelated.

Identification and estimation.—Models I, III, and their combination are identifiable by the arguments presented above. Model III can be identified directly from the parameters of single-equation estimates of the determinants of parental encouragement and college plans. Model I and the model that combines the effects of models I and III can be identified and estimated from the parameters of the equation for parental encouragement, a reduced-form equation for college plans, and the disturbance covariance for the two reduced-form equations. Given the normality assumption of the errors in the structural equations, the models were estimated by maximum likelihood probit analysis (see Appendix).[18]

Results.—Table 2 presents the log likelihood statistics for four models estimated by maximum likelihood probit analysis. From these models can be derived the structural parameters for the models of interest. The four include a model in which (1) there is no effect of parental encouragement

TABLE 2

LOG LIKELIHOODS AND LIKELIHOOD RATIO χ^2 TESTS FOR ALTERNATIVE MODELS OF
EFFECTS OF SOCIAL BACKGROUND ON PARENTAL ENCOURAGEMENT
AND COLLEGE PLANS

A. LOG LIKELIHOODS

Model*	Log Likelihood
A. No effect of parental encouragement on college plans........	$-10,903.21$
B. Effect of observed dummy variable for high parental encouragement on college plans (model III)	$-10,078.14$
C. Effect of unobserved continuous variable for parental encouragement on college plans (model I).................	$-10,069.75$
D. Effect of both unobserved continuous and observed dummy variables on college plans (models I and III)...............	$-10,069.06$

B. LIKELIHOOD RATIOS

Model Comparison	Likelihood Ratio χ^2†	Degrees of Freedom	P
B vs. A	1,650.1	1	$<.001$
C vs. A	1,666.9	1	$<.001$
D vs. B	18.2	1	$<.001$
D vs. C	1.4	1	$.2<P<.3$

* All models include effect of sex, IQ, and parental socioeconomic status on both parental encouragement and college plans. See text for discussion of data and measurement.

† Likelihood ratio χ^2 statistics are computed as -2 times the differences of the appropriate log likelihoods.

[18] A computer program for estimating two reduced-form equations with binary dependent variables and correlated disturbances by maximum likelihood probit analysis is available from us.

on college plans, (2) the dummy variable for encouragement affects college plans but the reduced-form disturbance covariance is assumed to be zero (model III), (3) the dummy variable for encouragement does not affect college plans but the disturbance covariance is nonzero (model I), and (4) the disturbance covariance is nonzero and there is a dummy variable effect of encouragement on college plans (models I and III combined).

The log likelihoods in table 2 indicate the relative fits of alternative models for the parental encouragement effect on college plans. When the log likelihood of a model is subtracted from the log likelihood for a model within which the first model is nested, -2 times the difference is distributed as χ^2 with degrees of freedom (df) equaling the difference in numbers of parameters between the two models (e.g., Hogg and Craig 1970). The lower panel of table 2 contrasts several pairs of models. The first two lines show the improvement in fit of the model when the two types of parental encouragement effects on college plans are added separately to a model that includes only background effects on college plans and on parental encouragement. The dummy and continuous latent variable effects reduce χ^2 by 1,650 and 1,667, respectively, both highly significant improvements in fit. To see whether both types of effects are necessary, a model that combines the effects of models I and III is contrasted with the models that include either the dummy variable or the latent continuous variable for parental encouragement. These contrasts, shown in the last two lines of table 2, show that whereas the elimination of the continuous latent variable from the model significantly weakens the fit of the model ($\chi^2 = 18.2$ with 1 df), the exclusion of the dummy variable has no significant impact on the fit of the model ($\chi^2 = 1.4$ with 1 df). Thus, relative to the combined model, model I, which includes the effect of a latent continuous variable for encouragement and excludes the dummy variable, cannot be rejected, but model III, which includes the encouragement effect as a dummy variable and excludes the latent variable, is rejected.

Table 3 presents the reduced-form probit coefficients for the determinants of parental encouragement and college plans under alternative models. These coefficients are expressed in raw form under the assumption that the disturbance variances of the equations are unity. As indicated by the asymptotic normal statistics, both the effects of the dummy variable for parental encouragement in model III and the disturbance correlation, from which is derived the effect of the latent variable for encouragement in model I, are highly significant (see the third and fourth columns). When both the error correlation and the dummy variables are included in the model, as shown in the last column, the error correlation remains highly significant, but the dummy variable coefficient is negative and insignificant. This is consistent with the goodness-of-fit results reported in table

TABLE 3

REDUCED-FORM MAXIMUM LIKELIHOOD PROBIT PARAMETER ESTIMATES FOR EFFECTS OF
BACKGROUND FACTORS ON PARENTAL ENCOURAGEMENT AND COLLEGE PLANS

		DEPENDENT VARIABLE				
			College Plans			
				Corresponding Structural Model		
INDEPENDENT VARIABLE	PARENTAL ENCOURAGEMENT*	No Parental Encouragement Effect	III	I	I and III
Sex (female)	−.359	−.255	−.121	−.254	−.283
	(−13.3)	(−9.0)	(−3.9)	(9.0)	(7.7)
IQ	.334	.452	.370	.450	.447
	(24.1)	(30.6)	(23.0)	(30.8)	(26.6)
SES	.538	.489	.302	.485	.514
	(37.3)	(32.5)	(17.8)	(32.5)	(19.6)
Parental encouragement (observed dummy variable)	...	...	1.372	...	−.383
			(38.1)		(−1.1)
ρ (correlation of errors in reduced-form equations)	...	...	...	.695	.846
				(57.2)	(6.2)
Constant	.250	.441	1.360	.438	.127
	(12.1)	(21.6)	(39.2)	(21.6)	(.41)

NOTE.—Coefficients are unscaled raw probit estimates which need to be rescaled if error variances are not assumed to be unity. Numbers in parentheses are asymptotic normal statistics (Z-score).

* Parental encouragement coefficients are invariant under choice of model.

2. These results suggest that parental encouragement is better modeled as a latent continuous variable in its effects on college plans than as the discrete variable that is directly measured.

Table 4 presents the structural parameters that are derived from the coefficients in table 3 through the identification methods presented in Section II. The first column presents the parental encouragement equation, and the second, third, and fourth columns present the coefficients for the reduced form and models III and I, respectively.[19] To obtain the coefficients for the parental encouragement equation in table 4, the raw coefficients are rescaled under the assumption that the latent variable for encouragement has variance of unity. This requires an estimate of the error variance which is, using the results of Section II.1,

$$\sigma_{e_p}^2 = 1/[b_1^2 \text{ var } (X) + b_2^2 \text{ var } (Q) + b_3^2 \text{ var } (S) + 2b_1b_2 \text{ cov } (X,Q)$$

$$+ 2b_1b_3 \text{ cov } (X,S) + 2b_2b_3 \text{ cov } (Q,S) + 1],$$

TABLE 4

STRUCTURAL COEFFICIENTS FOR EFFECTS OF BACKGROUND FACTORS ON PARENTAL
ENCOURAGEMENT AND COLLEGE PLANS UNDER MODELS I AND III

		DEPENDENT VARIABLE*		
		College Plans		
			Model	
INDEPENDENT VARIABLE	PARENTAL ENCOURAGEMENT	Reduced Form	III	I
Sex (female)	−.290	−.203	−.083	−.004
	(−13.3)	(−9.0)	(−3.9)	(−0.2)
IQ	.269	.360	.254	.175
	(24.1)	(30.6)	(23.0)	(17.6)
SES	.434	.389	.208	.091
	(37.3)	(32.5)	(17.8)	(8.2)
Parental encouragement (observed dummy)	...	...	.942	...
			(38.1)	
Parental encouragement (unobserved continuous)	...	...	...	.686
				(57.2)
Constant	.202	.351	.934	.251
	(12.1)	(21.6)	(39.2)	(17.7)
Error variance	.650	.633	.472	.327

NOTE.—Numbers in parentheses are asymptotic normal statistics (Z-scores).

* Calculations assume that unmeasured continuous variables for parental encouragement and college plans have variances of unity.

[19] Structural coefficients for the model combining the effects of models I and III are similar to those for model I.

where β_1, β_2, and β_3 are the raw coefficients reported in the first column of table 3. The variances and covariances of the observed independent variables are as follows:

	X	Q	S	P
X	.250	−.006	−.008	−.031
Q		1.000	.278	.162
S			1.000	.214
P				.250

Thus

$$\sigma_{\psi}^2 = 1/[(-.359)^2(.250) + (.334)^2 + (.538)^2 + 2(-.359)(.334)(-.006)$$

$$+ 2(-.359)(.538)(-.008) + 2(.334)(.538)(.278) + 1]$$

$$= .650.$$

With this estimate in hand, the rescaled coefficients can be obtained directly. For example, the rescaled coefficient for sex is $(-.359)(\sqrt{.650}$ $= -.290$. The coefficients for IQ, socioeconomic status, and the constant are similarly calculated as the products of $\sigma_{\epsilon_{\psi}}$ and the unscaled coefficients in table 3. The structural coefficients and error variances for the reduced-form and model III college-plans equations are derived by the same method.

The derivation of the structural parameters for model I, shown in the last column of table 4, follows the method outlined in Section II.2. The reduced-form error covariance for the parental encouragement and college-plans equation is

$$\rho\sqrt{\sigma_{\eta_c}^2\sigma_{\eta_{\psi}}^2} ,$$

where ρ is the disturbance correlation and σ_{ψ}^2 and $\sigma_{\eta_c}^2$ are the error variances of the reduced-form equations for parental encouragement and college plans, respectively. Thus cov $(\eta_c,\epsilon_{\psi}) = .695\sqrt{(.650)(.633)} = .446$. Then the structural coefficient for the effect of the latent variable for parental encouragement on college plans is $.446/.650 = .686$ reported in the final column of table 4. The remaining structural coefficients for the social background effects in model I are calculated from the reduced-form equations using the formulas in Section II.2. The coefficient for sex, for example, is $(-.203) - (-.290)(.686) = -.004$. The IQ, SES, and constant coefficients are calculated similarly. Finally, the error variance for the college-plans equation under model I is $\sigma_{\epsilon_c}^2 = \sigma_{\eta_c}^2 - \alpha_3^2\sigma_{\psi}^2 = .633 - (.686)^2(.650) = .327$.

The structural coefficients for the college-plans equation differ sub-

stantially between models I and III. The coefficients of the social background factors are smaller in model I, indicating weaker net effects than in model III. The coefficients for the two parental encouragement variables are not comparable inasmuch as they are measured in different scales. In model III the coefficient indicates that, for a latent variable indexing intention to attend college, there is a .942 standard deviation (SD) difference between individuals coded as high and those coded as low on parental encouragement. Under model I, the coefficient indicates that a 1 SD change in parental encouragement yields a .686 SD change in the latent variable for college plans. Finally, the error variance for model I is approximately 30% smaller than for model III, indicating that the model in which parental encouragement is treated as continuous is more successful in explaining college plans.

Direct and indirect effects of social background.—Models I and III have different implications for the allocation of the effects of social background on college plans into direct effects and indirect effects through parental encouragement. The decomposition of direct and indirect effects under the two models is shown in table 5. The top panel of the table shows the effects in SD units of the latent continuous variable for college plans. The bottom panel expresses the direct and indirect components as proportions of the total (reduced-form) effects of social background on

TABLE 5

DECOMPOSITION OF EFFECTS OF SEX, IQ, AND SOCIOECONOMIC STATUS ON
COLLEGE PLANS UNDER ALTERNATIVE MODELS

	MODEL I			MODEL III		
VARIABLE	Total	Direct	Indirect via Parental Encouragement	Total	Direct	Indirect via Parental Encouragement
Absolute effects:						
Sex	−.203	−.004	−.199	−.192	−.083	−.109
IQ	.360	.175	.185	.345	.254	.100
SES	.389	.091	.298	.371	.208	.163
Relative effects:						
Sex	1.000	.020	.980	1.000	.432	.568
IQ	1.000	.486	.514	1.000	.726	.290
SES	1.000	.234	.766	1.000	.561	.439

NOTE.—College plans are measured in probit transformed scale of probabilities of planning to attend college, that is, $\phi^{-1}(p)$, where ϕ is the cumulative normal function and p is the probability of planning to attend college. To obtain decomposition of the actual probability of attending college, multiply each entry by $(1/\sqrt{2\pi}) \exp(-z^2/2)$, where

$$p = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(\frac{-u^2}{2}\right) du$$

for a selected value of p (Hanushek and Jackson 1977, p. 189). At $p = .327$, the overall sample proportion planning to attend college, $z = -.448$ and $(1/\sqrt{2\pi}) \exp(-z^2/2) = .361$. Relative effects, however, are invariant under this transformation.

college plans. For model I, the absolute effects are calculated using the basic methods of path analysis. The direct effects are simply the structural coefficients for the background variables reported in the last column of table 4. The indirect effects are the products of the social background coefficients in the parental encouragement equation and the structural parameter for encouragement in the model I college-plans equation. For model III, the direct effects are again the structural coefficients for the social background factors obtained under that model and reported in the third column of table 4. The indirect effects are the products of three factors: the coefficient for the independent variable in the encouragement equation, the coefficient for the dummy variable for encouragement in the college-plans equation, and a factor $\exp{(-z^2/2)}\sqrt{2\pi}$, where

$$p = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(\frac{-u^2}{2}\right) du$$

for a selected value of p. This last factor is simply the ordinate of the normal curve corresponding to the "area" under the curve measured by p. Taking p to be the sample proportion with "high" parental encouragement, $p = .5188$, $z = .0471$, and $\exp{(-z^2/2)}/\sqrt{2\pi} = .3985$. Then, for example, the indirect effect of sex on college plans is $(-.290)(.942)(.3985)$ $= -.1089$.

The table of relative effects shows that under model I, a much larger proportion of the social background influences on college plans occurs through differential parental encouragement than under model III. For example, according to model I, three-fourths of the effect of parental socioeconomic status on college plans can be attributed to the effects of SES on parental encouragement and the effect of the latter on college plans. Under model III, however, less than one-half of the SES effect on college plans is due to the intervening mechanisms of parental encouragement. These results reflect the more powerful influence of the continuous latent variable for encouragement relative to that of the observed dichotomous variable.

2. Fair-Employment-Practices Legislation and Relative Wages

Landes (1968) and Heckman (1976) examine the possible effects of state-level Fair-Employment-Practices Legislation (FEPL) on the relative wages of blacks and whites. The correlation between the existence of FEPL and the level of black wages relative to that of whites is positive, but such a correlation may arise through several mechanisms. The legislation itself may have an active effect by deterring employers who would otherwise discriminate against blacks in wages. Alternatively, the law itself may

have no effect; it may be symptomatic of other social and historical factors that lead states not only to pass progressive laws but also to treat the races equally in the labor market. Under this Durkheimian view of the law, the accumulated sentiments in a state, not their legal manifestations, affect the relative wages of blacks and whites. For the purposes of this discussion, this factor will be termed "progressive sentiment." A third hypothesis is that neither the law nor progressive social conditions but other characteristics of states correlated with the existence of FEPL explain the positive zero-order correlations between the existence of FEPL and the relative standing of blacks.

These competing views can be represented in models II and IV. Under model IV, FEPL affects relative wages directly. Under model II, an unmeasured variable, of which FEPL is an indicator, affects relative wages, but FEPL itself does not. This latent variable can be interpreted as degree of progressive sentiment. In these models, the relationship between sentiment and FEPL is stochastic; that is, despite the positive association between FEPL and progressive sentiment, some progressive states may lack FEPL and, conversely, states where sentiments favoring discrimination are widespread may have FEPL. Under both models II and IV, progressive sentiment intervenes causally between other social factors and the relative wage levels of blacks and whites. Model IV cannot be identified with only a single indicator of the unmeasured variable. Moreover, progressive sentiment is better modeled as having several indicators. Thus a second indicator of progressive sentiment, based on state voting behavior, is included in the models discussed below.

Data and models.—The data for this example derive from published sources for the 48 continental United States during the 1950s and 1960s.[20] The relative wages of blacks and whites are measured as the ratio of wages of black males to those of white males in 1959 (B). Progressive sentiment ($P*$) is indicated by two binary variables: (1) whether a state had a Fair-Employment-Practices Law before 1959 (F), and (2) whether a state gave less than 10% of its vote to George Wallace in the 1968 presidential election (W).[21] Progressive sentiment is assumed to arise from past and contemporary social conditions. To reflect this, two independent variables are assumed to affect sentiment, the percentage of the civilian labor force that was unionized in 1953 (N) and the percentage of the white work force with more than 12 years of schooling in 1959 (E). Both vari-

[20] For further documentation of the variables discussed below, see Landes (1968). The data are available from us.

[21] The 10% cutoff point on the Wallace vote is arbitrary. Two dichotomous indicators of progressive sentiment are used in this example to follow the general form of model IV as discussed in Section II. A better formulation, which could be estimated by simple extensions of the methods discussed here, would treat the Wallace vote as a continuous indicator of progressive sentiment.

ables are assumed to affect progressive sentiment positively and to increase the likelihood of FEPL and a low Wallace vote. As discussed thus far, the model is $P^* = \beta_0 + \beta_1 N + \beta_2 E + \epsilon_p$, where ϵ_p is a normally distributed disturbance and var $(P^*) = 1$. In addition, $F' = \lambda_F P^* + \epsilon_F$, and $W' = \lambda_W P^* + \epsilon_W$, where var $(F') =$ var $(W') = 1$; ϵ_p, ϵ_F, and ϵ_W are mutually uncorrelated; $F = 1$ if $F' \geq M_1$; $W = 1$ if $W' \geq M_2$; $F = 0$ if $F' < M_1$; $W = 0$ if $W' < M_2$; and M_1 and M_2 are unobserved thresholds. This part of the model is similar to the relationships among X, Y^*, Y_1, Y_2, d_{y_1}, and d_{y_2} in figure 7.

In addition to FEPL and progressive sentiment, several other variables are assumed to affect relative wages. These include the ratio of nonwhite to white males in the civilian labor force in 1959 (L); the ratio of the percentage of the nonwhite male population living in urban areas to the corresponding percentage for white males in 1959 (R); the percentage of the total male population living in urban areas in 1959 (U); and the ratio of nonwhite male to white male mean years of schooling completed in 1959 (S). Then a general equation for the relative wages can be written $B = \alpha_0 + \alpha_1 L + \alpha_2 R + \alpha_3 U + \alpha_4 S + \alpha_5 F + \alpha_6 P^* + \epsilon_B$, where ϵ_B is a normally distributed disturbance that is uncorrelated with ϵ_p, ϵ_F, and ϵ_B. Under model II, $\alpha_5 = 0$, whereas under model IV, $\alpha_6 = 0$. If neither FEPL nor the sentiment that it indicates affects relative wages, then $\alpha_5 = \alpha_6 = 0$.[22]

Identification and estimation.—These models are identified using the methods discussed in Section II.5 for model IV. The measurement and structural equations for progressive sentiment can be identified separately from the relative wage equation. The models estimated here are overidentified inasmuch as each endogenous variable is affected by several exogenous variables that do not affect each other. Because no computer software that imposes the overidentifying restrictions is available, the estimates reported below are not unique but instead depend on the equations used in the solution. Moreover, since the sentiment and wage equations are estimated separately, no global goodness-of-fit statistics for the models are available. The R^2s of the alternative relative wage equations are compared in the discussion below, but they do not provide valid tests of the fits of the models.

Results.—Table 6 contains the reduced-form estimates for the equations predicting the two indicators of progressive sentiment. These estimates

---

[22] The rationale for these independent variables is discussed in Landes (1968). Heckman (1976) investigates the possible effects of FEPL and of progressive sentiment using somewhat different models from those investigated here. His models treat FEPL as related deterministically to the latent variable for sentiment; in this they are similar to models I and III. Heckman also explores, however, the possible simultaneity of legislation and relative wages, an issue not considered here.

are scaled such that their corresponding latent variables F' and W' have variances of unity. From these parameters the measurement and structural parameters for progressive sentiment are obtained. Thus, from the formulas given in Section II.5, and arbitrarily using the coefficient on schooling, measurement parameters are given by

$$\lambda_F = \sqrt{1 - .572 + \frac{(.140)}{(.141)}(.620)(.756)(.856)} = .909 \, ,$$

$$\lambda_W = \sqrt{1 - .731 + \frac{(.141)}{(.140)}(.620)(.756)(.856)} = .820 \, ,$$

and

$$\sigma_{\epsilon_F}{}^2 = 1 - (.909)^2 = .173 \, ; \quad \sigma_{\epsilon_W}{}^2 = 1 - (.820)^2 = .328 \, ,$$

$$\sigma_{\epsilon_F}{}^2 = \frac{(.620)(.756)(.856)}{(.909)(.820)} = .538 \, .$$

Given the measurement parameters, consistent estimates of the structural equations for P^* can be obtained by arbitrarily focusing on the parameter for FEPL: $\beta_0 = -5.897/.909 = -6.487$; $\beta_1 = .109/.909 = .120$; $\beta_2 = .140/.909 = .154$. These are reported in the first column of table 7.

TABLE 6

REDUCED-FORM MAXIMUM LIKELIHOOD PROBIT PARAMETER ESTIMATES FOR EFFECT
OF LEVELS OF SCHOOLING AND UNIONIZATION ON INDICATORS
OF PROGRESSIVE SENTIMENT

	DEPENDENT VARIABLES	
INDEPENDENT VARIABLES*	FEPL (F)	Wallace Vote (W)
Unionization (N)	.101	.056
	(3.3)	(2.1)
Schooling (E)	.140	.141
	(1.3)	(1.3)
Constant	−5.897	−3.716
	(−2.5)	(−1.8)
Error variance	.572	.731
Disturbance correlation	.620	
	(2.4)	
Log likelihood	−48.39	
N	48	

NOTE.—Coefficients are scaled under assumption that latent variables for FEPL and Wallace vote have variance of unity. FEPL denotes whether or not a state had a Fair-Employment-Practices Law before 1959. Wallace vote denotes whether or not a state gave less than 10% of its popular vote to George Wallace in the 1968 presidential election. Numbers in parentheses are asymptotic normal statistics (Z-scores).

* For discussion of independent variables, see text.

The equations for relative wages in which neither F nor $P*$ has an effect and in which only F has an effect are estimated directly by OLS. The results are reported in the second and fourth columns of table 7. The equations for relative wages that include an effect for progressive sentiment and that include both sentiment and FEPL effects are estimated as follows. First, estimated values of $P*$ are obtained using the structural parameters reported in the first column of table 7, that is, $\hat{P}* = -6.487 + .120N + .154E$ for each observation. Then the $\hat{P}*$s are entered into the relative wage equations in place of the unobserved $P*$. These equations are estimated by OLS and yield estimates for model II and the combined effects for models II and IV which are reported in the third and fifth columns of table 7.[23]

TABLE 7

STRUCTURAL COEFFICIENTS FOR EFFECTS OF SOCIAL FACTORS ON PROGRESSIVE
SENTIMENT AND RELATIVE WAGES UNDER MODELS II AND IV

		DEPENDENT VARIABLES			
		Relative Wages (B)			
			Model		
INDEPENDENT VARIABLES*	PROGRESSIVE SENTIMENT	No Effect of Law	II	IV	II and IV Combined
Unionization (N)	.120	...	...	...	...
Schooling (E)	.154	...	...	...	...
Relative labor force participation (L)	...	−.651 (.126)	−.563 (.115)	−.688 (.115)	−.608 (.110)
Relative schooling (S)	...	.115 (.146)	.081 (.130)	−.010 (.139)	−.009 (.128)
Relative urbanization (R)	...	.099 (.032)	.095 (.029)	.108 (.030)	.103 (.027)
Urbanization (U)	...	−.033 (.062)	−.134 (.062)	−.164 (.070)	−.216 (.067)
FEPL (F)	...	...	...	.073 (.023)	.056 (.022)
Progressive sentiment ($P*$) ...	...	...	.026 (.008)	...	.022 (.007)
Constant	−6.487	.524 (.128)	.628 (.118)	.678 (.127)	.723 (.118)
Error variance	.545	.165	.128	.134	.111
R^2†	...	.749	.806	.797	.832

NOTE.—Numbers in parentheses are estimates of coefficient standard errors under the assumption of OLS. For model II and the combination of models II and IV, the assumption of homoscedasticity is not met and thus estimated standard errors are not unbiased (Heckman 1978).

* For discussion of independent variables, see text.

† R^2 is calculated under assumptions of OLS.

[23] This procedure is equivalent to the method of solving overidentified models from reduced forms given in Section II.3.

Although these calculations preclude rigorous tests of significance, they suggest that both FEPL and progressive sentiment affected the relative economic standing of blacks in 1959. With neither effect present, the remaining variables explain approximately 75% of the variance in relative wages, whereas FEPL and progressive sentiment each add approximately 5% to explained variance when added separately, and together add 8% explained variance. The results suggest that simply looking at the effect of FEPL alone gives an overestimate of its impact: when progressive sentiment is controlled, the coefficient for FEPL drops by almost 25%, although it remains substantial.

V. CONCLUSION

This article has reviewed alternative formulations of the role of discrete variables in recursive structural equation models. Discrete variables arise in many ways in empirical social science, and structural equation models that are sensitive to these varying substantive contexts are available. In particular, discrete variables may enter structural equations as measures of inherently discrete processes or as indicators of unmeasured continuous variables. When discrete variables are dependent, they are best modeled as having underlying continuous counterparts, but these latent variables may be related either deterministically or stochastically to the observed discrete variables. This article has illustrated the alternative substantive contexts in which the various combinations of discrete variable characteristics may be appropriate. In addition, it has discussed a set of recursive models in which discrete and continuous dependent variables can be considered within single models. Thus, models with discrete variables have the same flexibility as structural equation models with only continuous dependent variables. Finally, this article has shown that methods of path analysis for continuous variables can be applied with little additional difficulty to systems of variables in which some endogenous variables are discrete.

APPENDIX

Methods of Estimation

As shown in Section II, model identification can be achieved typically through solving the structural equations for reduced-form equations, obtaining reduced-form estimates, and using these estimates to derive the structural parameters. For model III this step is unnecessary; structural estimates are estimable directly without resort to the reduced form. Models

I, II, and IV, however, in the absence of other overidentifying restrictions, require not only reduced-form coefficients but also an estimate of the covariances between the disturbances in the two equations of the model. The following discussion reviews procedures for estimating the reduced-form parameters and the disturbance covariance which yield the structural parameters through the steps outlined above. It first reviews methods for estimating single equations with binary response variables and then discusses methods for jointly estimating pairs of equations in which one or both of the dependent variables is discrete. These latter methods provide estimates of covariances among disturbances required for identifying models I, II, and IV.[24]

I. SINGLE-EQUATION ESTIMATION

As is well-known, logit and probit models estimated by maximum likelihood are suitable methods for estimating the effects of independent variables on a dichotomous dependent variable (Cox 1970; Finney 1971; Hanushek and Jackson 1977; Nerlove and Press 1976). These methods are briefly reviewed here. Recall equation (2), for the effects of an independent variable on a latent continuous variable Y^* of which d_y is an observed dichotomous indicator:

$$Y^*_i = \beta + \beta X_i + \epsilon_{yi},\qquad\qquad (2)$$

where i denotes the ith observation ($i = 1, \ldots, N$) and all other notation is as defined above. Let $c_{yi} = (\beta_0/\sigma_{\epsilon_y}) + (\beta_i/\sigma_{\epsilon_y})X_i$. Then

$$p(d_{yi} = 1) = \int_{-\infty}^{c_{yi}} f\left(\frac{\epsilon_y}{\sigma_{\epsilon_y}}\right)d\epsilon_y,$$

[24] This discussion emphasizes methods that are computationally feasible at the present time. There is no reliable general computer program for maximum likelihood estimation of structural equations with discrete variables that parallels LISREL for continuous variables (Jöreskog and Sörbom 1978; Jöreskog 1973). The methods for calculating reduced-form equations discussed below include both maximum likelihood and less efficient methods. Maximum likelihood estimation of reduced-form equations, however, does not, in general, yield maximum likelihood estimates of corresponding structural parameters inasmuch as overidentifying restrictions on the structural form of the model are not always imposed and may lead to multiple solutions for the same structural parameter. An alternative strategy, not used here, is to compute tetrachoric correlation coefficients between pairs of discrete variables and biserial correlations between discrete and continuous variables and to estimate structural equation models based on the resulting correlation matrix using LISREL (Jöreskog and Sörbom 1981). This method provides unique estimates of coefficients in overidentified models if the estimated correlation matrix is positive definite. The latter condition is not guaranteed by pairwise estimation of tetrachoric and biserial correlations, nor does this procedure yield correct estimates of coefficient standard errors or of test statistics.

$$= \int_{-\infty}^{c_{yi}} f(t)dt \ , \ $$

and

$$p(d_{yi} = 0) = \int_{c_{yi}}^{\infty} f(t)dt \ , $$

where f is the probability density function for the distribution followed by ϵ_{yi} and $t = \epsilon_y/\sigma_{\epsilon_y}$. Then if f is the density function of the extreme value distribution,

$$p(d_{yi} = 1) = \int_{-\infty}^{c_{yi}} \exp \ (t)/[1 + \exp \ (t)]^2 dt$$

$$= \exp \left(\frac{\beta_0}{\sigma_{\epsilon_y}} + \frac{\beta_1}{\sigma_{\epsilon_y}} X_i \right) \bigg/ \left[1 + \exp \left(\frac{\beta_0}{\sigma_{\epsilon_y}} + \frac{\beta_1}{\sigma_{\epsilon_y}} X_i \right) \right] \ , \tag{A1}$$

and

$$p(d_{yi} = 0) = 1 \bigg/ \left[1 + \exp \left(\frac{\beta_0}{\sigma_{\epsilon_y}} + \frac{\beta_1}{\sigma_{\epsilon_y}} X_i \right) \right] \ ,$$

which defines the logistic model. Then the likelihood is

$$L = \prod_{i=1}^{N} \left\{ \exp \left(\frac{\beta_0}{\sigma_{\epsilon_y}} + \frac{\beta_1}{\sigma_{\epsilon_y}} X_i \right) d_y \bigg/ \left[1 + \exp \left(\frac{\beta_0}{\sigma_{\epsilon_y}} + \frac{\beta_1}{\sigma_{\epsilon_y}} X_i \right) \right] \right\} \ ,$$

and maximum likelihood estimates are obtained by picking values of $\beta_0/\sigma_{\epsilon_y}$ and $\beta_1/\sigma_{\epsilon_y}$ that make L as large as possible (Cox 1970; Hanushek and Jackson 1977).

Alternatively, if ϵ_y follows a normal distribution,

$$p(d_{yi} = 1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{c_{yi}} \exp \left(\frac{-t^2}{2} \right) dt \ , $$

$$p(d_{yi} = 0) = \frac{1}{\sqrt{2\pi}} \int_{c_{yi}}^{\infty} \exp \left(\frac{-t^2}{2} \right) dt \ , \tag{A2}$$

which define the probit model. Then

$$L = \prod_{i}^{N} \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{c_{yi}} \exp \ (-t^2) dt \right]^{d_{yi}} \left[\frac{1}{\sqrt{2\pi}} \int_{c_{yi}}^{\infty} \exp \ (-t^2) dt \right]^{(1-d_{yi})} \ ,$$

and maximum likelihood estimates are obtained by picking values of β_0/σ_{e_y} and β_1/σ_{e_y} that make L as large as possible (Finney 1971; Hanu- shek and Jackson 1977). Computer programs for estimating logit and probit models by maximum likelihood are widely available (e.g., Baker and Nelder 1978).

II. LOGIT VERSUS PROBIT

For the estimation of single-equation models with dichotomous dependent variables, logit and probit models are virtually interchangeable. In spe- cialized circumstances, mathematical models of scientific phenomena may imply one or the other model (Berkson 1951), but in most applications, logit and probit models differ only in their underlying distributional as- sumption. This is a trivial difference given the similarity of the logistic and cumulative normal functions (Hanushek and Jackson 1977). Logit models are attractive because they can be derived from log-linear models for frequency data (Fienberg 1980) and provide a closed-form expression for probabilities (compare [A1] and [A2]), but this feature provides no advantage or disadvantage in the analysis of binary data per se.

For multiequation models, in contrast, the choice of logit or probit models is more consequential. As noted above, the main complication in estimating multiequation structural models with discrete endogenous vari- ables is the estimation of the covariance of the disturbances in the reduced- form equations. When the logit model is extended to more than a single dichotomous dependent variable, in its usual form it fails to yield an estimate of the disturbance covariance. In the multinomial logit model, the most common extension of the binary logit model in use, the dis- turbance covariance is assumed to be zero (McFadden 1974). Naturally, this precludes estimation of the cross-equation dependence implied by the models discussed here.[25] In contrast, the probit model is more easily generalized to a multiequation problem. For a pair of probit equations, the disturbances are bivariate normal, and their covariance can be esti- mated from the parameter for correlation in the bivariate normal model. Moreover, the assumption of normally distributed disturbances in the probit model facilitates estimation of models with both discrete and con- tinuous dependent variables. Models for the latter typically assume nor- mality of disturbances, and thus two-equation models with one continuous

---

[25] There are other bivariate distributions than the multinomial that possess logistic marginals and allow disturbances to be correlated. Gumbel (1961) discusses two such distributions, but they have the disadvantage that the correlation is either restricted to a specific value or varies over a limited range. McFadden (1978, 1980) considers a generalization of the extreme value distribution that allows for correlated errors, but this distribution has yet to achieve wide acceptance.

and one discrete response variable can be estimated under the assumption of bivariate normality of the disturbances when the discrete equation is specified as a probit model. In the following discussion, therefore, the multivariate normal model, based on extensions of probit analysis, is employed to estimate the reduced-form error covariance required for the identification of the models discussed above.[26]

The following discussion of the multivariate probit model first presents the likelihood equations for two equations with discrete dependent variables and for two equations when one has a discrete and the other a continuous dependent variable. Then it turns to alternative methods for estimating the coefficients and disturbance covariance for such models when computer software for maximum likelihood estimation for multi-equation systems is unavailable.

III. MAXIMUM LIKELIHOOD FOR TWO-EQUATION MODELS

1. Two Discrete Dependent Variables

Recall that the identification of model I is based on two equations, the structural equation for Y^* and the reduced-form equation for Z:

$$Y^*_i = \beta_0 + \beta_1 X_i + \epsilon_{yi} , \tag{11}$$

$$Z_i = \gamma_0 + \gamma_1 X_i + \eta_{zi} , \tag{12}$$

where i denotes the ith observation. The maximum likelihood procedure can be illustrated for this reduced-form model, but the identical estimation problem arises for model II (for the reduced forms for Y' and Z) and for model IV (for reduced forms for Y'_1 and Y'_2). Let Z be a latent continuous variable that is indicated by an observed dichotomous variable d_z. Assume that ϵ_y and γ_z follow a bivariate normal distribution with var $(\epsilon_y) = \sigma_{\epsilon_y}^2$, var $(\eta_z) = \sigma_{\eta_z}^2$, and cov $(\epsilon_y, \eta_z) = \rho\sigma_{\epsilon_y}\sigma_{\eta_z}$. Define $t_y = \epsilon_y/\sigma_{\epsilon_y}$, and $t_z = \sigma_{\eta_z}$ as standardized normal variables. Then the probability density function for t_y and t_z is the bivariate normal density (e.g., Hogg and Craig 1970):

$$g(t_y, t_z) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-(t_y^2 - 2\rho t_y t_z + t_z^2)/(1 - \rho^2)\right] . \tag{A3}$$

---

[26] This argument implies that when the disturbance covariance is unnecessary to achieve identification, as in the modification of model I in which there is an instrumental variable that affects Y^* but not Z, the logit model may still be used. For an example of this approach, see Duncan and Duncan (1978, pp. 287–96). Nonetheless, when Z is continuous, it remains attractive to estimate both equations under the assumption of normally distributed errors.

For this model there are four possible outcomes on the dependent variables, namely (1) $d_y = d_z = 1$; (2) $d_y = 1$, $d_z = 0$; (3) $d_y = 0$, $d_z = 1$; (4) $d_y = d_z = 0$. Let $c_{yi} = (\beta_0/\sigma_{\epsilon_y}) + (\beta_1 X_i/\sigma_{\epsilon_y})$ and $c_{zi} = (\gamma_0/\sigma_{\eta_z}) + (\gamma_1 X_i/\sigma_{\eta_z})$. Given the bivariate normal model, then

$$p(d_{yi} = 1, d_{zi} = 1) = \int_{-\infty}^{c_{yi}} \int_{-\infty}^{c_{zi}} g(t_{yi}, t_{zi}) dt_z dt_y ,$$

$$p(d_{yi} = 1, d_{zi} = 0) = \int_{-\infty}^{c_{yi}} \int_{c_{zi}}^{\infty} g(t_{yi}, t_{zi}) dt_z dt_y ,$$

$$p(d_{yi} = 0, d_{zi} = 1) = \int_{c_{yi}}^{\infty} \int_{-\infty}^{c_{zi}} g(t_{yi}, t_{zi}) dt_z dt_y ,$$

$$p(d_{yi} = 0, d_{zi} = 0) = \int_{c_{yi}}^{\infty} \int_{c_{zi}}^{\infty} g(t_{yi}, t_{zi}) dt_z dt_y ;$$

and the likelihood is

$$L = \prod_{i=1}^{N} [p(d_{yi} = d_{zi} = 1)]^{d_{yi} d_{zi}} [p(d_{yi} = 1, d_{zi} = 0)]^{d_{yi}(1 - d_{zi})}$$

$$[p(d_{yi} = 0, d_{zi} = 1)]^{d_{zi}(1 - d_{yi})} [p(d_{yi} = d_{zi} = 0)]^{(1 - d_{yi})(1 - d_{zi})} .$$

Maximum likelihood estimates are obtained by picking values of $\beta_0/\sigma_{\epsilon_y}$, $\beta_1/\sigma_{\epsilon_y}$, γ_0/σ_{η_z}, γ_1/σ_{η_z}, and ρ that make L as large as possible. The error variances $\sigma_{\epsilon_y}^2$ and $\sigma_{\eta_z}^2$ can be derived through scale restrictions on Y^* and Z, and thus the disturbance covariance can be obtained from the equation

$$\text{cov} (\eta_z, \epsilon_y) = \rho \sigma_{\epsilon_y} \sigma_{\epsilon_z} . \tag{A4}$$

2. One Continuous and One Discrete Dependent Variable

Now consider model I again but allow Z to be an observed continuous variable instead of a latent variable. Identical estimation problems arise for model II (for the reduced forms for Y' and Z). As before, define $t_y = \epsilon_y/\sigma_{\epsilon_y}$, but define $t_z = (Z - \gamma_0 - \gamma_1 X)/\sigma_{\eta_z}$, which indicates that Z, unlike Y^*, is observed. Again, let $g(t_y, t_z)$ be the bivariate normal density defined above (A3). Then the likelihood is

$$L = \prod_{i=1}^{N} \left[\int_{-\infty}^{c_{yi}} g(t_y, t_z) dt_y \right]^{d_y} \left[\int_{c_{yi}}^{\infty} g(t_y, t_z) dt_y \right]^{(1 - d_y)} .$$

Maximum likelihood estimates are obtained by picking values of β_0/σ_u, β_1/σ_u, γ_0, γ_1, σ_m, and ρ that make L as large as possible. The error variance σ_u^2 can be derived through a scale restriction on Y^* (no such restriction is required to identify σ_m^2 since Z is observed). As before, the disturbance covariance is obtained from (A4).

IV. NONLINEAR WEIGHTED LEAST SQUARES ESTIMATORS

The methods just discussed provide efficient estimates of the reduced-form parameters and the disturbance covariances, from which the structural parameters may be derived. These methods are computationally expensive and difficult to implement in the absence of reliable computer software. Most critically, when the number of discrete endogenous variables exceeds two or three, the computation of multivariate normal probabilities becomes intractable inasmuch as each iteration of calculation requires the evaluation for each observation of a probability which is a multiple integral of the same order as the number of discrete endogenous variables (e.g., Muthén 1979). Fortunately, alternative methods are available that hold the promise of allowing consistent (though not efficient) estimation of overidentified structural equation models with discrete variables as well as correct estimation of parameter standard errors and suitable test statistics. These methods, noted briefly here, are discussed more fully elsewhere (Avery, Hansen, and Hotz 1981; Avery and Hotz 1981).

The alternative estimation methods, which employ a nonlinear least squares procedure, use only a subset of the information available about the joint distribution of the discrete endogenous variables that is used in maximum likelihood. In particular, they use the univariate and bivariate moments of the discrete variables but ignore the higher-order moments, thereby avoiding the calculation of higher-order multiple integrals. The estimation strategy is as follows. First, solve the structural equation model for its reduced form and consider the nonlinear functions for each dichotomous dependent variable and for each pair of dependent variables, that is, for the jth dichotomous endogenous variable d_{yj} ($j = 1, \ldots, p$) (excluding a subscript for individual observations),

$$d_{yj} = \int_{-\infty}^{\sum_m \Pi_{mj} X_m} \phi(t_j) dt_j + v_j ,$$

and, defining $d_{ykl} = d_{yk} d_{yl} (l < k \leq p)$,

$$d_{ykl} = \int_{-\infty}^{\sum_m \Pi_{mk} X_m} \int_{-\infty}^{\sum_m \Pi_{ml} X_m} g(t_k t_l) dt_k dt_l + w_{kl} ,$$

where d_y is a dichotomous variable taking the value 1 or 0, Π_{mj} is the reduced-form parameter for the effect of the mth exogenous variable X_m on the jth endogenous variable d_{yj}, $\phi(t_j)$ and $g(t_k t_l)$ are the standard univariate and bivariate normal probability density functions, respectively, and v_j and w_{kl} are residuals.

Second, apply a weighted nonlinear least squares criterion to estimate the structural parameters from the system of $p + p(p - 1)/2$ reduced-form equations by minimizing the weighted sums of squares of the residuals v_j and w_{kl} with respect to the structural parameters. The weighting procedure takes account of the differential variances associated with different sums of squared residuals and provides a rigorous method of combining overidentifying restrictions to yield unique estimates. The minimization and the solution of the resulting system of nonlinear equations can be carried out using standard methods for nonlinear estimation (e.g., Goldfeld and Quandt 1972). The procedure leads straightforwardly to estimates of the variances and covariances of the estimated parameters (Avery and Hotz 1981). Although these methods are not yet in wide use, they have been applied in special cases (Avery et al. 1981). Computer software soon to be available will allow their application to the general set of models discussed in this article (Hotz, personal communication, 1982).

REFERENCES

Alwin, Duane F., and Robert M. Hauser. 1975. "The Decomposition of Effects in Path Analysis." *American Sociological Review* 40 (February): 37–47.

Amemiya, Takeshi. 1974. "Bivariate Probit Analysis: Minimum Chi-Square Methods." *Journal of the American Statistical Association* 69 (December): 940–44.

———. 1976. "The Maximum Likelihood, the Minimum Chi-Square and the Nonlinear Weighted Least-Squares Estimator in the General Qualitative Response Model." *Journal of the American Statistical Association* 71 (June): 347–51.

———. 1978. "The Estimation of a Simultaneous Equations Generalized Probit Model." *Econometrica* 46 (September): 1193–1206.

———. 1981. "Qualitative Response Models: A Survey." *Journal of Economic Literature* 19 (December): 1483–1536.

Amemiya, Takeshi, and Frederick Nold. 1975. "A Modified Logit Model." *Review of Economics and Statistics* 57:255–57.

Ashford, J. R., and R. R. Sowden. 1970. "Multi-variate Probit Analysis." *Biometrics* 26 (September): 535–46.

Avery, R. B., L. P. Hansen, and V. J. Hotz. 1981. "Multiperiod Probit Models and Orthogonality Condition Estimation." Discussion Paper 81-12. Chicago: Economics Research Center, National Opinion Research Center, University of Chicago.

Avery, R. B., and V. J. Hotz. 1981. "Estimation of Multiple Indicator Multiple Cause (Mimic) Models with Dichotomous Indicators." Working Paper 60-80-81. Pittsburgh: Graduate School of Industrial Administration, Carnegie-Mellon University.

Baker, R. J., and J. A. Nelder. 1978. *The GLIM System. Release 3. Generalised Linear Interactive Modelling.* Oxford: Royal Statistical Society.

Berkson, Joseph. 1951. "Why I Prefer Logits to Probits." *Biometrics* 7 (December): 327–39.

Bielby, William T., and Robert M. Hauser. 1977. "Structural Equation Models." *Annual Review of Sociology* 3:137–61.

Brier, S. 1978. "The Utility of Systems of Simultaneous Logistic Response Equations." Pp. 119–29 in *Sociological Methodology 1979,* edited by Karl F. Schuessler. San Francisco: Jossey-Bass.

Clogg, Clifford. 1980. "New Developments in Latent Structure Analysis." Pp. 215–46 in *Factor Analysis and Measurement in Sociological Research,* edited by D. J. Jackson and E. F. Borgatta. Beverly Hills, Calif.: Sage.

Cox, D. R. 1970. *The Analysis of Binary Data.* London: Methuen.

Davis, James. 1975. "Analyzing Contingency Tables with Linear Flow Graphs: D Systems." Pp. 111–45 in Heise, ed., 1975.

Davis, James, and Susan R. Schooler. 1974. "Non-Parametric Path Analysis—Multivariate Structure of Dichotomous Data When Using the Odds Ratio or Yule's Q." *Social Science Research* 3:267–97.

Duncan, Beverly, and Otis D. Duncan. 1978. *Sex Typing and Social Roles.* New York: Academic Press.

Duncan, Otis D. 1966. "Path Analysis: Sociological Examples." *American Journal of Sociology* 72:1–16.

———. 1975. *Introduction to Structural Equation Models.* New York: Academic Press.

Duncan, Otis D., David L. Featherman, and Beverly Duncan. 1972. *Socioeconomic Background and Achievement.* New York: Seminar.

Fienberg, Stephen E. 1975. "Comment." *Journal of the American Statistical Association* 70 (September): 521–23.

———. 1980. *The Analysis of Cross-classified Categorical Data.* 2d ed. Cambridge, Mass.: MIT Press.

Finney, D. J. 1971. *Probit Analysis.* 3d ed. Cambridge: Cambridge University Press.

Goldberger, Arthur S. 1971. "Econometrics and Psychometrics: A Survey of Communalities." *Psychometrika* 36:83–107.

Goldberger, Arthur S., and Otis D. Duncan, eds. 1973. *Structural Equation Models in the Social Sciences.* New York: Seminar.

Goldfeld, Stephen M., and Richard E. Quandt. 1972. *Nonlinear Methods in Econometrics.* Amsterdam: North-Holland.

Goodman, Leo A. 1972. "A General Model for the Analysis of Surveys." *American Journal of Sociology* 77 (May): 1035–86.

———. 1973a. "The Analysis of Multidimensional Contingency Tables When Some Variables Are Posterior to Others: A Modified Path Analysis Approach." *Biometrika* 60:179–92.

———. 1973b. "Causal Analysis of Data from Panel Studies and Other Kinds of Surveys." *American Journal of Sociology* 78 (March): 1135–91.

———. 1974. "The Analysis of Systems of Qualitative Variables When Some of the Variables Are Unobservable. Part I—a Modified Latent Structure Approach." *American Journal of Sociology* 79 (March): 1179–1259.

———. 1979. "A Brief Guide to the Causal Analysis of Data from Surveys." *American Journal of Sociology* 84 (March): 1078–95.

———. 1981. "Association Models and the Bivariate Normal for Contingency Tables with Ordered Categories." *Biometrika* 68:347–55.

Gortmaker, Steven L. 1979. "Poverty and Infant Mortality in the United States." *American Sociological Review* 44 (April): 280–97.

Gumbel, E. J. 1961. "Bivariate Logistic Distributions." *American Statistical Association Journal* 56 (June): 335–49.

Hanushek, Eric A., and John E. Jackson. 1977. *Statistical Methods for Social Scientists.* New York: Academic Press.

Hausman, J., and D. Wise. 1978. "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences." *Econometrica* 46 (March): 403–26.

Heckman, James. 1976. "Simultaneous Equation Models with Both Continuous and Dis-

crete Endogenous Variables with and without Structural Shift in the Equations." Pp. 235–72 in *Studies in Nonlinear Estimation,* edited by Stephén M. Goldfeld and Richard E. Quandt. Cambridge, Mass.: Ballinger.

———. 1978. "Dummy Endogenous Variables in a Simultaneous Equation System." *Econometrica* 46 (July): 931–59.

Heckman, James, and Burton Singer. 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica.* In press.

Heise, David R., ed. 1975. *Sociological Methodology 1976.* San Francisco: Jossey-Bass.

Hogg, Robert V., and Allen T. Craig. 1970. *Introduction to Mathematical Statistics.* 3d ed. London: Macmillan.

Jencks, Christopher, et al. 1979. *Who Gets Ahead? The Determinants of Success in America.* New York: Basic.

Johnson, Norman, and Samuel Kotz. 1970. *Continuous Univariate Distributions-1.* New York: Wiley.

Jöreskog, Karl. 1973. "A General Method for Estimating a Linear Structural Equation System." Pp. 85–112 in Goldberger and Duncan, eds., 1973.

Jöreskog, Karl, and Dag Sörbom. 1978. *LISREL IV User's Guide.* Chicago: National Educational Resources.

———. 1981. *LISREL. Analysis of Linear Structural Relationships by the Method of Maximum Likelihood. Version V.* Uppsala: University of Uppsala.

Landes, William. 1968. "The Economics of Fair Employment Laws." *Journal of Political Economy* 76 (July/August): 507–52.

Lazarsfeld, P. F., and N. W. Henry. 1968. *Latent Structure Analysis.* Boston: Houghton Mifflin.

Leik, Robert. 1975. "Causal Models with Nominal and Ordinal Data: Retrospective." Pp. 271–75 in Heise, ed., 1975.

McFadden, Daniel. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." Pp. 105–42 in *Frontiers in Econometrics,* edited by Paul Zarembka. New York: Academic Press.

———. 1976 "Quantal Choice Analysis: A Survey." *Annals of Economic and Social Measurement* 5 (Fall): 363–90.

———. 1978. "Modelling the Choice of Residential Location." *Transportation Research Record* 673:72–77.

———. 1980. "Econometric Models of Probabilistic Choice." Pp. 198–272 in *Structural Analysis of Discrete Data,* edited by Charles F. Manski and Daniel McFadden. Cambridge, Mass.: MIT Press.

McKelvey, Richard D., and William Zavoina. 1975. "A Statistical Model for the Analysis of Ordinal Level Dependent Variables." *Journal of Mathematical Sociology* 4:103–20.

Manski, Charles F. 1981. "Structural Models for Discrete Data: The Analysis of Discrete Choice." Pp. 58–109 in *Sociological Methodology 1981,* edited by Samuel Leinhardt. San Francisco: Jossey-Bass.

Muthén, Bengt. 1979. "A Structural Probit Model with Latent Variables." *Journal of the American Statistical Association* 74 (December): 807–11.

Nerlove, Marc, and S. James Press. 1976. "Multivariate Log-Linear Probability Models for the Analysis of Qualitative Data." Discussion Paper no. 1. Chicago: Center for Statistics and Probability, Northwestern University.

Pearson, K. 1900. "Mathematical Contributions to the Theory of Evolution in the Inheritance of Characters Not Capable of Exact Quantitative Measurement, VIII." *Philosophical Transactions of the Royal Society,* ser. A, 195:79–150.

Rosenthal, Howard. 1980. "The Limitations of Log-linear Analysis." *Contemporary Sociology* 9 (March): 207–12.

Sewell, William H., and Robert M. Hauser. 1975. *Education, Occupation, and Earnings.* New York: Academic Press.

Sewell, William H., and V. P. Shah. 1968. "Social Class, Parental Encouragement, and Educational Aspirations." *American Journal of Sociology* 73:559–72.

Stolzenberg, Ross M. 1979. "The Measurement and Decomposition of Causal Effects in Nonlinear and Nonadditive Models." Pp. 459–88 in *Sociological Methodology 1980*, edited by Karl F. Schuessler. San Francisco: Jossey-Bass.

Yule, G. U. 1900. "On the Association of Attributes in Statistics: With Illustration from the Material of the Childhood Society." *Philosophical Transactions of the Royal Society*, ser. A, 194:257–319.

Zellner, Arnold, and Tong Hun Lee. 1965. "Joint Estimation of Relationships involving Discrete Random Variables." *Econometrica* 33 (April): 382–94.

The Hierarchy of the Sciences?[1]

Stephen Cole
State University of New York at Stony Brook

For 200 years it has been assumed that the sciences are arranged in a hierarchy, with developed natural sciences like physics at the top and social sciences like sociology at the bottom. Sciences at the top of the hierarchy presumably display higher levels of consensus and more rapid rates of advancement than those at the bottom. A distinction is made between two classes of knowledge: the *core,* or fully evaluated and universally accepted ideas which serve as the starting points for graduate education, and the *research frontier,* or all research currently being conducted. Data are presented from a set of empirical studies which show that, at the research frontier, there are no systematic differences between sciences at the top and at the bottom of the hierarchy in either cognitive consensus or the rate at which new ideas are incorporated. It is concluded that in all sciences knowledge at the research frontier is a loosely woven web characterized by substantial levels of disagreement and difficulty in determining which contributions will turn out to be significant. Even at the research frontier, however, minimal levels of consensus are a necessary condition for the accumulation of knowledge. Consensus in all sciences is maintained by sociological processes such as the evaluation and reward systems.

It is commonly assumed by sociologists, natural scientists, and educated laymen that there are substantial differences between the natural and the social sciences. For seven years I have been conducting empirical studies in which I have compared various aspects of both the cognitive and the social structures of the natural and the social sciences. This paper presents a theoretical integration and interpretation of all the data collected thus far. I begin by examining the commonly accepted views about the ways in which knowledge in the natural sciences differs from that in the social sciences.

THE HIERARCHY HYPOTHESIS

Almost 200 years ago, Auguste Comte, the father of the history of science, set out what he called the "hierarchy of the sciences," maintaining that the sciences progress through ordained stages of development at quite different rates. Thus, for Comte, astronomy, the most general of all the sciences, develops first and is followed successively by physics, chemistry, biology, and finally sociology. The hierarchy of the sciences described not only the complexity of the phenomena studied by the different sciences but also their stage of intellectual development.[2]

Contemporary discussions of knowledge in the different sciences do not differ significantly from that of Comte. A generation ago, James Conant (1950) proposed that the various branches of science differ in "degree of empiricism." Norman Storer (1967) makes a distinction between the so-called hard natural sciences and the soft social sciences. Thomas Kuhn (1970) differentiates among the sciences by the extent to which they have a developed paradigm or shared theoretical structures and methodological approaches about which there is a high level of consensus. The sciences at the bottom of the hierarchy (e.g., sociology) are assumed to exhibit less consensus and are frequently referred to as being in a "preparadigmatic" state.

Drawing on recent thought in this tradition, Zuckerman and Merton (1973a) proposed that the cognitive texture of scientific fields differs in the extent to which they are codified. Although they do not define the concept in detail, codification refers to the extent to which a field's paradigm or theoretical orientation is systematically developed. "Codification refers to the consolidation of empirical knowledge into succinct and interdependent theoretical formulations" (p. 507). Physics and biochemistry are hypothesized to be relatively codified fields; geology and zoology are less so; sociology, anthropology, and political science are presumably among the least codified. They further suggested that the degree of codification of knowledge would be reflected in the level of agreement on the significance of specific scientific contributions and on the importance of work of particular scientists. Variables which underlie the concept of the hierarchy of sciences are indicated in table 1.

For the past 200 years, it has been assumed that there are differences among the sciences in levels of cognitive consensus. Highly codified fields like physics are assumed to have substantially higher levels of agreement than are less codified fields like sociology. Before presenting some of the empirical results of the studies I have conducted that were designed to test this hypothesis, it is necessary to make a conceptual distinction be-

[2] Of course, for Comte—unlike his followers—sociology was at the apex of the hierarchy rather than at the bottom.

tween the two classes of knowledge that are studied. Historians, philosophers, and sociologists of science have tended to treat knowledge as if it were a uniform whole. There are some obvious and clear-cut differences in the types of knowledge found in fields like physics and sociology. For example, in physics there are a small number of theories which are universally accepted. The cornerstone of contemporary physics is the theory of quantum mechanics. Few modern physicists reject the validity and significance of quantum mechanics. Likewise, if we consider such a specialty as superconductivity research, virtually all physicists working in this specialty recognize the 1957 theory of Bardeen, Cooper, and Shrieffer as the foundation of all contemporary work in this field. At least on the surface, there seem to be no comparable examples in a field like sociology.

Because we have these obvious differences between the two fields, sociologists have assumed that we would find similar differences in all areas of knowledge. I would like to suggest, however, that knowledge is not a uniform whole, that it is made up of two principal components that I shall call the *core* of knowledge and the *research frontier*. The core consists of a small set of theories and analytic techniques which represent the "given" at any particular point in time. Thus, if we were to look at

TABLE 1

CHARACTERISTICS OF DIFFERENT TYPES OF SCIENCE

Variable	Top of the Hierarchy	Bottom of the Hierarchy
1. Development of theory	Highly developed theory; research guided by a paradigm; high levels of codification	No theory or low-level generalization; preparadigmatic phase; low levels of codification
2. Quantification.........	Ideas expressed in mathematical language	Ideas expressed in words
3. Cognitive consensus....	High levels of consensus on theory, methods, significance of problems, significance of an individual's contribution	Low levels of consensus on theory, methods, significance of problems, significance of an individual's contribution
4. Predictability..........	Ability to use theory to make verifiable predictions	Inability to make verifiable predictions
5. Rate of obsolescence ...	High proportion of references to recent work as older work becomes obsolete—indicating significant cumulation of knowledge	Low proportion of references to recent work as older work remains just as important as new work—indicating lack of significant cumulation of knowledge
6. Rate of growth	"Progress" or the rate at which new knowledge grows is relatively fast	"Progress" or the rate at which new knowledge grows is relatively slow

the content of courses taught advanced undergraduate students or first-year graduate students in a field like physics, we would acquire a good idea of the content of the core of knowledge. The core is the starting point, that knowledge which people take as a given from which new knowledge will be produced. The core is characterized by having a relatively small number of theories and substantial consensus on the importance of these theories.

The other component of knowledge, the research frontier, consists of all the work currently being done by all active researchers in a given discipline. The research frontier is where all new knowledge is produced. As figure 1 illustrates, the research frontier is linked to the core through the evaluation process. Most new knowledge, however, turns out to be of little or no lasting significance (Cole and Cole 1972). As time progresses, a small number of ideas continue to be used and assessed as important. Some of these eventually become part of the core.[3]

Since historians and philosophers of science naturally are interested in the most important knowledge, most of their theoretical formulations have been based upon a study of the core. And since our ideas about the
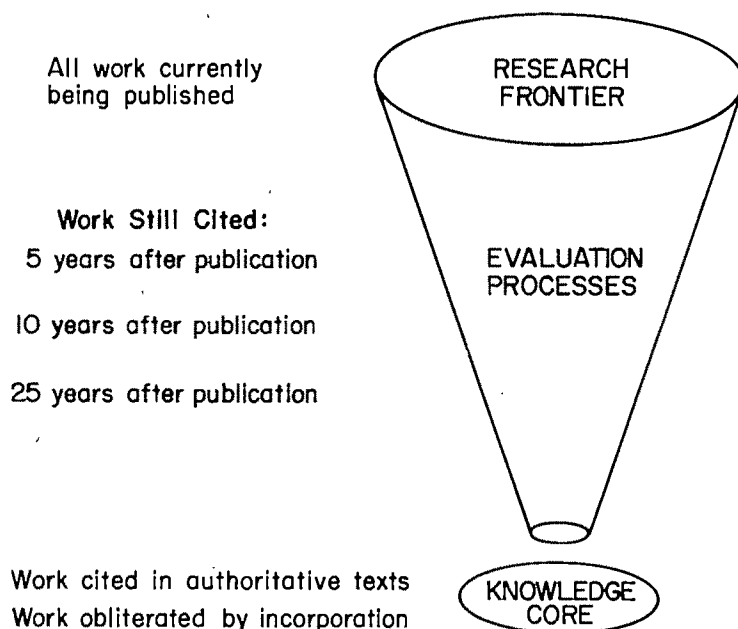

All work currently being published

Work Still Cited:

5 years after publication

10 years after publication

25 years after publication

Work cited in authoritative texts
Work obliterated by incorporation

RESEARCH FRONTIER

EVALUATION PROCESSES

KNOWLEDGE CORE

FIG. 1.—The structure of knowledge in science

[3] Clearly, some ideas move from the frontier to the core faster than others. An analysis of the social processes affecting the evaluation of ideas over time should increase our understanding of how consensus is developed and maintained.

structure of knowledge in various scientific fields have been heavily influenced by a reading of both historians and philosophers of science, we have assumed that the research frontier will be similar in structure to the core. As a sociologist, I am particularly interested in studying the social processes influencing the behavior of scientists as they work at the frontier. The studies described below represent attempts to measure the level of cognitive consensus at the research frontier in various scientific disciplines. Below I shall present some data from a preliminary study that suggest that there is substantially greater consensus in the core of the natural sciences than in that of the social sciences. The question is, Are these same differences found at the research frontier?

Since sociologists have tended to assume that there are substantial differences in the way in which knowledge is organized in the various fields, very little work has been done on this problem. An exception is the 1972 paper by Lodahl and Gordon. Of particular interest was their survey of the perceptions among scientists in four fields (physics, chemistry, sociology, and political science) about the relative degree of paradigm development of seven scientific fields, including their own. After a brief explanatory introduction, they asked: "Please rank the following fields on the degree to which you feel there is consensus over paradigms (law, theory and methodology) within the field . . ." (p. 59). The results conform exactly to the hierarchy that most of us would intuit. Physics had the highest level of paradigm development, followed by chemistry, biology, economics, psychology, political science, and sociology (with some disagreement about the order of the last two). This finding is an important one—not so much as a demonstration that these fields actually do rank order in levels of paradigm development in this way, but as the best evidence to date that scientists in all fields strongly *believe* that a hierarchy exists and that the perceptions of the hierarchy among scientists in different fields are almost identical.

In a study of evaluation patterns in science, Zuckerman and Merton (1973b) observed significantly lower rejection rates for journals in the natural sciences than for those in the social sciences. This was taken as evidence of higher levels of consensus in the natural sciences.[4] In a qualitative study of chemists and sociologists who were denied tenure in Ph.D.-granting departments, Rubin (1975) found chemists more likely to blame themselves and sociologists to dispute the validity of the criteria used to evaluate them. This difference is attributed to assumed differences in the level of paradigm development and consensus in the two fields. In one of the most systematic studies of field differences, Hargens (1975)

---

[4] Others, like Hargens (1975), have pointed out that differences in rejection rates may be due in part to noncognitive factors such as the amount of space available in the journals in the various fields.

presents data on chemistry, mathematics, and political science. He makes a distinction between normative integration (cognitive consensus) and functional integration, or the degree to which scientists in a field are familiar with each other's work and interact with each other in a community. Using the data then available, he concluded that political science was less normatively integrated than the two other fields, but that mathematics was the least functionally integrated.

Most studies dealing with field differences in consensus have either assumed the differences to exist, frequently using Lodahl and Gordon for authority, or used indirect data such as journal rejection rates.

EMPIRICAL STUDIES

For the past seven years I have been conducting a series of empirical studies aimed at measuring field differences. They have all focused on either levels of cognitive consensus or rate of obsolescence. At first I adopted a research strategy of assuming that there would be differences in the level of cognitive consensus in various fields and that these differences would in turn create measurable differences in other variables. The first study compared the reward system of five different fields. I hypothesized that the level of cognitive consensus in a field would have a significant influence on the structure of its reward system. In fields at the top of the hierarchy, like physics and chemistry, there would be substantial agreement on who is doing important work, and the sheer quantity of publications would therefore have a relatively small independent effect on the distribution of rewards. In fields at the bottom of the hierarchy, like psychology and sociology, there should be less agreement on who is doing important work. In the absence of such agreement, the sheer quantity of publications would have a greater influence on distribution of rewards. In fields with low levels of consensus, it may be relatively difficult to determine whether a paper is an important or unimportant contribution, but it is relatively easy to count the number of papers.

In order to test this hypothesis, I collected data on scientists who were promoted from associate to full professors at Ph.D.-granting institutions between 1964 and 1969 (Cole 1978).[5] The fields included were biochemistry, chemistry, physics, psychology, and sociology. The dependent variables for the study were three different forms of academic recognition. The two independent variables were the quantity of publications and the quality of work as measured by the number of citations received.[6]

[5] Space does not permit a full description of all the empirical studies. Such descriptions are available in the cited publications.

[6] Extensive past research indicates that citations are a valid indicator of the relative quality or impact of work. The number of citations is highly correlated with all other measures of

For all five fields, the quality of work was more strongly correlated with each of the three dependent variables than was quantity. In the 15 comparisons made, there was not one in which quantity showed a higher correlation than quality. In all five fields, the amount of variance explained by quantity after quality was accounted for was insignificant. And the joint effects of the two variables on, for example, visibility (the proportion of respondents familiar with a scientist's work) varied from a low of .56 for psychology to a high of .66 for physics. Using a regression model representing the reward system of science, I was able to show that there were no significant differences among the five fields.[7] The hypothesis that differences in levels of cognitive consensus among fields would create differences in their reward systems turned out to be incorrect.[8]

Several different studies, similar in logic to the one on scientific reward systems, were conducted on the effects of age on scientific output in six different fields: chemistry, geology, mathematics, physics, psychology, and sociology (Cole 1979). Zuckerman and Merton (1973a) have hypothesized that, in fields at the top of the hierarchy, it should be easier for young scientists to make significant discoveries than in fields at the bottom of the hierarchy. The primary reason for this is the method of gaining competence in fields of varying degrees of codification. In the highly codified fields, knowledge is compacted. Graduate students can quickly learn the current state of their field from textbooks and, while still students, begin work on the research frontier (Kuhn 1970). In the less codified fields, knowledge is not as compacted, and far greater experience is needed to gain competence: "In these [the less codified fields], scientists must get command of a mass of descriptive facts and of low-level theories whose implications are not well understood" (Zuckerman and Merton 1973a, p. 507). Zuckerman and Merton conclude: "Codification facilitates mastery

quality that sociologists of science have employed. As long as we keep in mind that research of high quality is being defined as research that other scientists find useful in their current work, citations provide a satisfactory indicator. Citations do not measure the "absolute" or "objective" quality of research, but they do measure the currently assessed value of work to colleagues who are themselves doing research.

[7] I did this by using a sample in which I combined data from all five fields after having standardized all the variables separately within fields and then used field as a dummy variable in a set of interaction terms. The amount of additional variance explained by including the interaction terms was negligible, leading to the conclusion that all five fields have similar reward systems (Cole 1978, pp. 182–83).

[8] Any theoretical argument depending on the acceptance of the null hypothesis, in this case that there are no systematic differences in the levels of consensus among fields, requires particular care with regard to type II errors. Such errors may result from sampling variability, measurement error, improper specification of models, and many other problems that occur in empirical research. Although in each study discussed I made every possible effort to avoid type II errors, each study can undoubtedly be faulted. The conclusions of this paper, however, depend not on the results of any specific study but rather on the pattern of results from all the studies. I thank an *AJS* reviewer for pointing this out.

of a field by linking basic ideas in the theoretical framework and by reducing the volume of factual information that is required to do significant research. This should lead scientists in the more codified fields to qualify earlier for work at the research front . . ." (p. 510).

Among the studies conducted in order to test this hypothesis empirically was one set concerning random samples of scientists currently employed in Ph.D.-granting institutions in the six fields under consideration. I obtained data on the number of papers each scientist had published in a recent five-year period and the number of citations received by these recent papers. The number of citations received was used as an indicator of the quality of current scientific output of the scientists of varying ages. The studies showed, contrary to expectation, that younger people were not more likely to publish more and receive more citations than their older colleagues. In fact, the relationship between both quantity and quality of publications and age is slightly curvilinear. It tends to rise through the thirties and peak in the late thirties or early forties and then taper off in the fifties and sixties (Cole 1979).

Of primary relevance here is the fact that young people in the more codified fields were no more likely to make significant discoveries than were young people in the less codified fields. The proportion of scientists under the age of 35 whose work published between 1965 and 1969 received more than the mean number of citations for their particular field is reported in table 2. Geology is the field in which the highest proportion of young scientists have made significant discoveries, and mathematics is the field in which the lowest proportion of young scientists have made significant discoveries. Furthermore, young psychologists are just about as likely to make important discoveries as young physicists and chemists. Other studies (also reported in Cole [1979]) examined the age at which the most eminent members of these disciplines made their first important

TABLE 2

PROPORTION OF SCIENTISTS UNDER 35 WHOSE
WORK RECEIVED MORE THAN THE MEAN NUMBER OF
CITATIONS FOR THEIR FIELD

Field	% Receiving More than Field Mean	N Scientists
Chemistry	29	115
Geology	38	45
Mathematics	11	101
Physics	25	138
Psychology	24	151
Sociology	13	60

discovery. It turned out that there were not significant differences in the age of first important discovery among the various fields. In fact, if we consider "professional age" (the number of years since obtaining Ph.D.) rather than chronological age, we find that there is no meaningful difference between the age at which a physicist makes his or her first important discovery and the age at which a sociologist makes his or her first important discovery.

The studies of scientific reward systems and of the age at which young scientists make important contributions were similar in the sense that neither contained a direct measure of the level of cognitive consensus or the level of codification. Rather, the methodology employed assumed that there were differences in these variables that would influence measured variables. In neither case did I find the hypothesized differences. This, of course, did not lead to the conclusion that there were no differences in levels of consensus or in levels of codification. Other possible explanations for the failure to find the hypothesized relations could be measurement error or the possibility that, despite differences in levels of cognitive consensus and codification, these variables would not influence those under study—the reward system and the relationship between age and scientific output. The next set of studies I conducted involved attempts to measure directly the level of cognitive consensus in various scientific fields.

I was able to use data from the comparative reward system study to provide several direct measures of the level of cognitive consensus in various scientific fields. As part of that study we asked random samples of scientists employed in Ph.D.-granting institutions to evaluate the work of 60 of their colleagues (Cole, Cole, and Dietrich 1978). For each name, the raters were asked to respond to the following: "Please indicate the relative importance of the work of the following scientists: has made very important contributions, has made above-average contributions, has made average contributions, work has been relatively unimportant, unfamiliar with work but have heard of scientist, have never heard of this scientist."

Responses obtained from such a questionnaire item have several evident limitations. One is that the question asks respondents to evaluate the significance of a colleague's work without taking into account the extent to which they are actually familiar with that work; some of the scientists answering the question may have only a hazy, superficial idea of a rated scientist's work, while others may have detailed familiarity with it. We do not know whether the degree and extent of this type of knowledge vary from field to field.

Acknowledging the limitations of the indicator, I computed the standard deviations of the ratings received by each of the 60 scientists. The smaller the standard deviation, the greater the agreement in evaluation.

I then computed the mean standard deviation for each of the five fields. As table 3 shows, differences among fields followed the expected pattern, with physics displaying the greatest agreement on evaluation and sociology the least.The differences, however, were small, and only that between physics and sociology was statistically significant at the .05 level.

Another set of data enabled us to measure this form of consensus in a second way. On the same questionnaires, we asked the raters to list the five scientists who had contributed most to their discipline in the past two decades. Presumably, scientists in highly codified fields would show greater agreement in this choice than those in less codified fields. The results were tabulated in two different ways: the percentage of total mentions going to the five most frequently named scientists and the number of different scientists mentioned divided by the total number of mentions. The higher the first proportion and the lower the second, the greater the approximation to consensus. The results, reported in table 4, again suggest that there are no large or systematic differences in this type of consensus between fields at the top and the bottom of the hierarchy. In fact, if we

TABLE 3

CONSENSUS ON EVALUATING SCIENTISTS BY FIELD
(60 Scientists in Each Field)

Field	N Raters	Mean Standard Deviation of Ratings
Biochemistry.	107	.71
Chemistry	111	.69
Physics.	96	.63
Psychology	182	.74
Sociology.	145	.76

TABLE 4

CONSENSUS ON WHICH SCIENTISTS HAVE CONTRIBUTED
THE MOST IN PAST TWO DECADES BY FIELD

Field	Mentions Received by 5 Most Mentioned Names (%)	Mentions Going to Discrete Names* (%)	N Mentions
Biochemistry.	41	21	473
Chemistry	34	39	377
Physics.	47	23	375
Psychology	32	36	435
Sociology.	36	19	661

* Number of different scientists mentioned divided by the total number of mentions.

use the percentage of mentions going to discrete names as a measure, sociology has the highest level of agreement, and chemistry the lowest. It should be pointed out again that the indicator of consensus used here may measure only impressionistic images of scientists' work rather than concrete knowledge of it.

Another, more concrete, measure of consensus is the distribution of citations in a scientific journal. In fields characterized by a high level of intellectual agreement, we would expect to find a heavy concentration of references to a relatively small number of papers and authors. The distribution of references in fields characterized by lack of agreement should more closely approximate a random distribution. Using the Gini coefficient as a measure of concentration (the higher the coefficient, the greater the concentration), we analyzed the distribution of citations in 1971 in 108 leading scientific journals (Cole, Cole, and Dietrich 1978).[9] Using the journal article as the unit of analysis, we computed a Gini coefficient for each journal. Examining the Gini coefficients, we do indeed find significant differences in the concentration of citations in the various journals. Data on the range and mean found in journals in the various fields are presented in table 5. Relatively small differences are found in the mean Gini coefficients of the various fields. The two social sciences, psychology and sociology, have about the same mean Gini coefficients as geology and mathematics. Examination of the complete range of data shows that, for example, the psychology journal with the highest Gini coefficient, *Journal of the Experimental Analysis of Behavior,* had a higher coefficient than any journal in chemistry, geology, or mathematics. Clearly, considerable differences in the coefficients occur for different journals in the same field. For instance, in chemistry, *Analytical Chemistry* has a Gini coefficient of .06, whereas the *Journal of Chemical Physics* has a coefficient of .27.

---

[*] Although the Gini coefficient has been used as a measure of consensus in science, there is a significant conceptual problem in its use. The Gini coefficient simultaneously measures two concepts. One is consensus, and the other is the dispersion of recognition in a particular field. It would, e.g., be possible to have complete consensus in a field about the importance of group of scientists' work, and yet all the scientists could be assessed as being equally important, i.e., receiving the same number of citations. Thus, if one has 100 scientists rating the work of 10 other scientists, all 100 raters might agree that the work of each of the 10 rated scientists was equivalent in value. Under such a circumstance the Gini coefficient would be zero but consensus would be complete. If the 100 raters had complete consensus but decided that there were wide differences in the significance of the contributions of the 10 rated scientists, the Gini coefficient would be high. In practice it is impossible to distinguish the extent to which a Gini coefficient is influenced by the dispersion of evaluations or consensus among scientists on the significance of work. In defense of the use of the Gini coefficient in comparative studies such as this one, it could be claimed that there is now significant evidence that the distribution of recognition in all scientific fields is highly skewed. If dispersion in all scientific fields is approximately the same, differences in the Gini coefficient should measure differences in consensus. Because of the difficulty in actually proving this, however, studies using the Gini coefficient should be seen as suggestive, not conclusive.

Similar variances are found in other fields that we intuitively believe to have a high level of codification; also, large differences in coefficients are obtained for journals in fields that we believe to have a lower level of codification. A good example is psychology, with scores varying between .05 and .29. When the two social sciences were taken as one group and all other fields as another, an analysis of variance showed that the differences were not statistically significant.

The last set of data on cognitive consensus in various scientific fields to be used here comes from a study of the peer review system employed by the National Science Foundation (NSF) in deciding which scientists should receive NSF research grants (Cole and Cole 1980; Cole, Rubin, and Cole 1978). We collected data on applications made to the NSF in fiscal year 1975 for 10 different scientific programs: algebra, anthropology, biochemistry, chemical dynamics, ecology, economics, fluid mechanics, geophysics, meteorology, and solid-state physics. For each of these programs we had information on approximately 100 applicants, about half of whom received NSF grants.

The NSF peer review system is based primarily on mail reviews obtained by a set of ad hoc mail reviewers. When a proposal is submitted, the program director selects four or five (in some cases more) reviewers who he or she believes are knowledgeable in the area. The proposal is sent to the reviewers, who are told to evaluate it on the basis of its scientific merit and the ability of the principal investigator. Raters are asked to give the proposal a score ranging from excellent (50) to poor (10). The ratings given proposals by these mail reviewers have a very strong effect on the final NSF decision. In fact, in most programs the mean rating obtained by the proposal explains in a probit regression analysis more than 75% of the variance on decision (Cole, Rubin, and Cole 1978). For the purposes of this paper, I am interested in the extent to which reviewers in various scientific fields agree on the merit of the proposals sent to them

TABLE 5

CONCENTRATION OF CITATIONS TO RESEARCH ARTICLES
IN SELECTED FIELDS

FIELD	N JOURNALS	GINI COEFFICIENTS	
		Mean	Range
Biochemistry...........	10	.21	.05–.34
Chemistry.............	12	.15	.06–.27
Geology..............	7	.10	.04–.23
Mathematics..........	6	.09	.06–.13
Physics..............	10	.18	.06–.35
Psychology	8	.16	.05–.29
Sociology.............	7	.09	.05–.11

to review. A proposal is a very concrete set of plans for research. Evaluating the proposal involves deciding whether the problem outlined is a significant one, determining whether the specific plans to investigate it are feasible, and evaluating the competence of the principal investigator. If there is indeed a higher level of cognitive consensus in the more codified fields at the top of the hierarchy than in those at the bottom of the hierarchy, we would expect to see greater consensus among reviewers of research proposals in the highly codified than in the less codified fields.

This hypothesis was tested in two different ways. First, I computed a one-way analysis of variance in which the dependent variable was the individual reviewer's rating and the independent variable was the proposal identification number. The analysis of variance divided the total amount of variance in the ratings into two parts: that which could be attributed to within-proposal differences among reviewers and that which could be attributed to between-proposal differences. In all 10 fields, I found approximately 50% of the total variance on ratings to be a result of dissensus among the reviewers of a given proposal. The data are presented in table 6. The percentage of variance explained by within-proposal disagreement ranged from a low of 35% for economics to a high of 63% for meteorology. The two social sciences included in the study, economics and anthropology, had the lowest level of disagreement about proposals.

As another way of measuring consensus among reviewers, my collaborators and I used the standard deviation of the ratings for each proposal.

TABLE 6

CONSENSUS AMONG MAIL REVIEWERS OF SCIENTIFIC PROPOSALS
SUBMITTED TO THE NSF*

Program	Total Variance due to Within-Proposal Variance (%)	Mean Standard Deviation of Reviewers' Ratings	Mean of Reviewers' Ratings	Coefficient of Variation
Algebra..........	45	.31	2.1	.15
Anthropology.....	43	.59	2.5	.24
Biochemistry.....	49	.60	2.6	.23
Chemical dynamics......	45	.42	2.3	.18
Ecology	54	.69	2.3	.30
Economics.......	35	.34	2.6	.13
Fluid mechanics ..	43	.61	2.8	.22
Geophysics.......	57	.61	2.4	.25
Meteorology......	63	.69	2.8	.25
Solid-state physics	55	.35	2.2	.16

* Data for each field are based on 100 proposals.

We then computed the mean standard deviation for the 100 proposals in each field. This mean ranged from a low of .31 in algebra to a high of .69 in ecology and meteorology. In using the mean standard deviation as a measure of agreement, one must also take into account the mean rating of reviewers. Therefore we used the coefficient of variation (mean standard deviation divided by mean rating). The results obtained from this analysis confirm those obtained from the analysis of variance. Once again economics proved to be the field with the lowest level of disagreement.

The results of the peer reviewer consensus study were replicated in a recently completed quasi experiment (Cole and Cole 1981; Cole, Cole, and Simon 1981). For three NSF programs, chemical dynamics, solid-state physics, and economics, we took a sample of 50 proposals submitted to NSF in fiscal year 1976. Of the total, 25 were approved for funding and 25 were declined. Before the results of the NSF decision were made public, we were given copies of these proposals. Then, with the aid of the members of the National Academy of Sciences, we selected a new group of reviewers. Half of them reviewed a proposal identical with the one which the NSF evaluators had received, and the other half received a proposal from which we had removed information on the identity of the investigator. The data from this experiment are presented in table 7. First, I should point out that among the reviewers used by the National Science Foundation we once again find approximately 50% of the variance attributable to within-proposal disagreement among the mail reviewers. Once again economics has the lowest amount of within-proposal disagreement, 43%, and this time chemical dynamics has the highest amount, 60%. Among the two experimental groups again, approximately 50% of the variance in each case was explained by within-proposal disagreement. This was even true in the case of the blind proposals which the reviewers were instructed to evaluate strictly on the basis of their opinion of the

TABLE 7

PROPORTION OF TOTAL VARIATION IN REVIEWS ACCOUNTED
FOR BY WITHIN-PROPOSAL VARIATION

FIELD	NSF MAIL REVIEWERS	COSPUP EXPERIMENT	
		Nonblind	Blind
Chemical dynamics	.60	.53	.63
	(242)	(213)	(212)
Solid-state physics	.51	.49	.51
	(192)	(190)	(203)
Economics.	.43	.47	.48
	(163)	(182)	(199)

NOTE.—Numbers in parentheses are the total number of individual reviewers for the group of 50 proposals.

scientific merits. These data lead to the conclusion that there is not a significant difference in levels of agreement on research proposals for fields located at the top of the hierarchy and those located in a lower position in the hierarchy.

The last set of empirical studies I shall describe was aimed not at measuring cognitive consensus but at measuring the rate at which knowledge became obsolete. It is frequently claimed that the fields at the top of the hierarchy show faster rates of knowledge obsolescence than those at the bottom. The indicator most often used to support such claims is called the "immediacy effect," or the tendency of scientists to cite very recent work in their papers. In highly codified fields we should find a faster rate of what Merton (1968) has described as "obliteration by incorporation" than in fields with lower levels of codification. It is not that the research is no longer important but that it has been incorporated into the general body of knowledge and therefore is no longer cited explicitly.

Zuckerman and Merton (1973a) have suggested that the age of references in journal literature may be an appropriate indicator of codification: "The journals in fields we intuitively identify as more highly codified—physics, biophysics and chemistry—show a larger share of references to recent work; they exhibit a greater 'immediacy,' as Derek Price calls it" (p. 508). In the work of Kuhn and of Zuckerman and Merton, the suggestion is at least implicit that rapid incorporation and a corresponding high immediacy of citations is an indicator of the extent to which a science is growing in a cumulative fashion. The extent to which recent work is utilized in current research may thus be seen as an indicator of the presence of conditions necessary for rapid scientific advance.

A preliminary study by Zuckerman and Merton on the age distribution of journal references provided some data that support the use of this measure as an indicator of codification. They report data (some collected by Price and others by them) which seem to show a correlation between the degree of codification (intuitively determined) and the proportion of references to work published in a recent five-year period. However, the data in their paper present two problems. The less significant one is that the reported data are based on a small number of journals unsystematically selected, and the reported statistics are based on small samples of articles within each journal. The more significant problem is that no control is provided for the age distribution of the entire body of literature in each field.[10]

Suppose that, in a particular field, 50% of all the articles ever published have been published in the past five years. If, in this same field, 50% of

---

[10] Zuckerman and Merton cite both Price (1970) and MacRae (1969) on the necessity of making such controls.

all references in current literature are to work published in the past five years, there is no immediacy effect. For, if citations were drawn at random, the age distribution of the literature would lead one to expect that roughly 50% of them would refer to work published in the previous five years. That is, recent work would have no higher probability of being cited than older work, and the age of references would be simply a function of the growth rate of the research area. To determine accurately the immediacy effect, the distribution of references in the current literature must be compared with the age distribution of that literature.

To see whether controlling for the growth of the literature would yield results different from those reported by Zuckerman and Merton, my collaborators and I examined the age distribution of *all* the references in the same 108 journals for which we had computed Gini coefficients. To estimate roughly the age distribution of the relevant literature in each of the selected scientific fields, we used the number of abstracts listed in the appropriate abstracting journals.[11]

For table 8, the immediacy effects for the five-year references were computed as follows. Using an abstracting journal, we added up all journal articles listed between 1920 and 1970 and determined what percentage of them were published between 1966 and 1970. This number was subtracted from the proportion of references to work published in that five-year period. For example, in psychology, 23% of the 1920–70

TABLE 8

IMMEDIACY EFFECTS FOR SELECTED SCIENTIFIC FIELDS

Field	N Journals	Literature Published in Last Five Years of Study (1966–70) (%)	Average References to Work Published in 1966–70 (%)	Average Immediacy Effect
Biochemistry	10	32	62	30
Chemistry	12	30	55	25
Geology	7	27	48	21
Mathematics	6	31	45	14
Physics	10	38	57	19
Psychology	8	23	47	24
Sociology	7	*	40	*

* No reliable method of calculating the size of the literature was available.

[11] To do this we were forced to make the assumption that work published before 1920, which constitutes only a small fraction of all scientific work, is essentially irrelevant today. Clearly, we are not implying that pre-1920s scientific work is irrelevant in the sense that it has not been crucially necessary for scientific advance, but only that most of it has been fully incorporated and is, therefore, rarely cited explicitly. In fact, very few citations to work published before 1920 appeared in any of the journals.

literature was published in that period. Since 47% of the current citations refer to work that is five years old or less, the immediacy effect is 24.

The most difficult problem in computing immediacy effects is to obtain even approximate data on the growth rates of research areas. Ideally we should know the growth rate of the literature in each of the research areas. We were forced, however, to use data on the growth rates of literature in an entire field. Although the data on the size of the literature are clearly inadequate and must be refined by further research, they do not suggest that fields at the top of the hierarchy have larger immediacy effects than fields lower down. For example, the immediacy effects of psychology, which is presumably less codified than physics, chemistry, or biology, are essentially similar to the immediacy effects for those fields. Given these data, it is not surprising that an analysis of variance showed no statistically significant difference between psychology journals and natural science journals in immediacy effects.

The data in table 8 suggest that all scientific fields have a slight tendency to favor recent work in their citation patterns. Is this tendency a peculiar and distinguishing characteristic of science, or does the same pattern appear in the journal literature of the humanities? Derek Price's (1970) data showed substantially lower proportions of citations to recent work in humanities journals than in either the natural or the social science journals. But Price considered neither the growth rate of the relevant journal literature nor the differences between citations to "data sources" and citations to other work in the area. For instance, if a literary critic is writing an article on the 17th-century sermons of John Donne, some references will be to the works of John Donne (the data source) and others to the work of other literary critics or other sources. These two types of references must be distinguished before the immediacy effect for a journal in the humanities can be computed.

I have completed a small pilot study of journal citations in English literature journals. A random sample of 25 articles was drawn from the 1971 volume of the *Publications of the Modern Language Association* (*PMLA*) and another 25 articles from the 1971 volume of *Studies in English Literature*. The references were then separated into "data citations" and "influence citations" to provide the distinction mentioned above. The mean age of all references in the 50 articles from the two journals was 43 years. However, and unsurprisingly, the mean age of the 300 data citations was 83 years, and the mean age of the 463 influence citations was 18 years.

Nevertheless, even when we separate influence citations from data citations and control for the growth of the literature, which was determined from the number of references listed in the Modern Language Association's *International Bibliography* from 1927 to 1970, we find a

negative immediacy effect for both journals. Thus, although approximately 29% of the literature was published in the last five years covered by the study, only 13% of the influence citations in *Studies in English Literature* and 21% of the influence citations in *PMLA* referred to work published in that period. The negative immediacy effects were, respectively, −16 and −8. The data for the two journals are summarized in table 9.

Clearly, the citation practices of scholars in English literature differ from those of scientists. The former are likely to draw equally on work produced in the past and work produced recently. Although we must study the humanities disciplines in greater detail, the immediacy effect may enable us to distinguish between a literature that is scientific and one that is not, even if it may not allow us to distinguish between highly codified and less codified scientific fields.

INTERPRETATION OF FINDINGS

All the studies that I have completed have failed to confirm the hypothesis that we would find higher levels of cognitive consensus and a larger immediacy effect among the more highly developed sciences at the top of the hierarchy than we would among the less developed social sciences at the bottom of the hierarchy. This leaves us with a difficult interpretive problem. If the hypothesis had been confirmed, there would have been some evidence for the interpretation that the attributes of the knowledge body itself influence both the level of cognitive consensus and the rate of advance as measured by the immediacy effect—that the structure of knowledge at the research frontier mirrors that found in the core. Since we have not found field differences in the level of consensus or in the

TABLE 9

IMMEDIACY EFFECT IN TWO ENGLISH LITERATURE JOURNALS
(463 Influence Citations; Data Citations Excluded)

AGE (In Years before 1970)		*Studies in English Literature*		*PMLA*	
Of Cited Work	Percentage of All Literature* (%)	Influence Citations (%)	Immediacy Effect	Influence Citations (%)	Immediacy Effect
0–5	29	13	−16	21	−8
0–10	50	43	−7	51	0
0–15	66	69	3	65	−1
0–20	76	83	7	79	3

* References listed in MLA *International Bibliography* 1927–70. Linguistics excluded. Sections included were: American, English (language and literature), medieval and neo-Latin, general, and miscellaneous.

immediacy effect, we may conclude tentatively that these variables are not primarily determined by the attributes of the core of knowledge itself, the level of codification or development, or the type of phenomena being studied. However, we are still faced with the question of what does influence the level of consensus at the research frontier.

Methodological Problems

Perhaps the failure to confirm the hierarchy hypothesis can be traced to methodological inadequacies. Indeed, when I first began this research program, I believed that the failure to find empirical support for the hypothesized field differences was simply a matter of inadequate measurement. Perhaps I had failed to measure the relevant aspects of cognitive structure, and, once proper measures had been developed, the hypothesized differences would be found. I no longer believe that the failure to find systematic differences in consensus among the various scientific fields is simply a matter of measurement difficulty. Particularly convincing were the data from the peer review study. These data were not responses to questionnaires; they were based on the actual evaluations made by working scientists in the course of carrying out one of their professional responsibilities. The fact that as much disagreement occurred among reviewers in the natural sciences as occurred among those in the social sciences lends strong support to the prior evidence collected from surveys and the citation index.

The major methodological problem in conducting this type of research is selecting the proper unit of analysis. For most of the studies on which I have reported data, the unit of analysis has been the scientific field. But clearly, for many cases, the scientific field may be more meaningful as a sociological entity than as an intellectual one. Scientific fields are now divided into specialties. It is not infrequent for researchers working in one specialty to be unfamiliar with both the theories and methods currently being used in the other specialties. Thus, physicists working in high-energy physics are not familiar with work being done in solid-state physics. Many physiological psychologists, for example, have only limited familiarity with work being done in social psychology. Aggregating scientists who work in diverse specialties may reduce the level of consensus.

Although it is important to select the proper unit of analysis to conduct this type of research, it is unlikely that the results are an artifact of the units of analysis used in the previous research. This is especially true of the peer review data. Those data came from specific programs, not from fields. Thus, the data on chemistry were from the program of chemical dynamics, which is only one of seven subsections of the National Science Foundation chemistry section. On the other hand, the data from the social

science anthropology included proposals submitted by both social anthropologists and physical anthropologists. Therefore, the result for this study could not have been an artifact of aggregation difficulties.

One criticism of my interpretation of the peer review data could be that consensus is created by the program director who might send proposals only to reviewers who approach the problem from roughly the same point of view as the principal investigator. Thus it might be argued that, although there is greater potential for disagreement in the social sciences, this disagreement is avoided because reviewers are selected only from among those who approach the problem in the same way as the applicant. As an example, consider a proposal from a psychologist to study depression. We might expect that neo-Freudian clinical psychologists and experimental psychologists would have sharply differing views on how to approach this problem. If the proposal was from an experimentalist and reviewed only by experimentalists, the potential for disagreement would not show up.

This criticism brings up once again the question of what the unit of analysis should be. From a strictly cognitive point of view it might make sense to define as the unit of analysis all work being done on a particular cognitive problem. Indeed, a traditional intellectual historian might do this. Thus, all people studying depression or cancer causation might be grouped together even if they do not define themselves as being part of the same community and have little or no contact with one another or even awareness of each other's work. On the other hand, we might take as the unit of analysis a community of scientists who identify themselves as such and who interact and are familiar with each other's work. In the postscript to the second edition of *The Structure of Scientific Revolutions,* Kuhn argues that this ambiguity over the unit of analysis caused some of the confusion generated by his book and that his concern was with scientific communities: "Both normal science and revolutions are . . . community-based activities. To discover and analyze them, one must first unravel the changing community structure of the sciences over time. A paradigm governs, in the first instance, not a subject matter but rather a group of practitioners. Any study of paradigm-directed or of paradigm-shattering research must begin by locating the responsible group or groups" (pp. 179–80).

From a sociological point of view, a scientific community is a more appropriate unit of analysis than an artificially constructed group of people created by the researcher who judges them to be working on the same problem. Clinical psychologists and experimental psychologists form two different scientific communities, although they sometimes study the same subject matter. The conclusions I have drawn about levels of consensus

are meant to apply only to existing scientific communities. Even if the subject matter was used as the unit of analysis, empirical research would be required to see whether the potential for disagreement is in fact greater for topics in social sciences than for those in the natural sciences.

The Core and the Frontier

In explaining why I found few systematic differences in the levels of consensus among scientific fields, let me refer again to the concept of knowledge discussed above. Although "knowledge" has generally been treated by sociologists as a uniform whole, it may be a useful heuristic device to divide knowledge into two classes, the core and the research frontier. Our imagery of knowledge in the various sciences has been based on the writings of historians and philosophers who for the most part pay little attention to contemporary knowledge but study only that part of the frontier which has stood the test of time and become part of the core. The measures I have obtained in the studies discussed above all deal with knowledge currently being produced on the research frontier.

The structure of knowledge at the frontier is not similar to the structure of knowledge at the core. Whereas the core contains a relatively small number of theories or exemplars, the frontier contains a much broader and more loosely woven web of knowledge. By this I mean that science at the research frontier is a great deal less rational and predictable than our imagery suggests. Let us briefly review some evidence supporting this contention. First, I have shown that there are substantial levels of disagreement about who is doing important work, what are important problems, what research should be funded. As just one example of the substantial difficulty in evaluating work at the frontier, I present one finding from the NSF peer review experiment. As reported above, we had independently selected reviewers reevaluate 50 proposals in each of three NSF programs (Cole and Cole 1981). Since the funding decision is a dichotomous one, we were able to look at the number of decisions which would have been reversed if our reviewers had been used rather than the NSF reviewers. Given that half of the NSF proposals were funded and half refused, a new decision based on a coin flip would yield a reversal rate of 50%. We found that, of the 150 proposals, 29% would have been reversed if the second set of reviewers had made the decision. This leads to the conclusion that whether an NSF proposal is funded is about half the result of chance. A components-of-variance model showed that the observed reversal rates could be explained entirely by the high levels of disagreement among reviewers of the same proposal (Cole et al. 1981).

There is a growing body of additional evidence that casts doubt on the

ability of scientists to predict which ideas produced at the research frontier
will turn out to be the most important. Thus, Henry Small (1974) has
found a weak correlation between ratings given by referees to articles
and citations received once the articles were published in the journals.
Grace Carter (1974), in a study of the NIH peer review system, found
only a weak correlation between priority scores given to proposals by
review panels and later citations of work that emerges from these projects.
She also found that the priority score given a proposal when it is first
submitted was correlated moderately ($r = .40$) with a score given to the
same proposal when it was resubmitted for renewal. We have found very
similar results in our study of the NSF peer review system. Scores given
to proposals by peer review raters are only weakly correlated with num-
bers of papers and citations emerging from the grant. And in one subject
we studied, algebra, we found that in the five-year period following the
NSF's decision, people whose proposals were declined actually turned
out to be on average slightly more productive than those whose proposals
were accepted. Although there may be some discoveries which are im-
mediately judged to be highly significant and do indeed turn out to be
highly significant, for most science it is not clear to participants, when
work is produced, whether it will turn out to be important. Although we
have always assumed that this is true for the social sciences, the studies
cited above suggest that it is also true for the natural sciences.

   Up to this point we have assumed that the differences hypothesized to
exist between the natural and the social sciences by Kuhn, Zuckerman
and Merton, and many others would apply to the core of these disciplines.
In an attempt to provide some data on this topic, I have conducted a
preliminary investigation of undergraduate textbooks in chemistry, phys-
ics, and sociology. If we assume that textbooks represent the content of
the core of a discipline, we may examine references in a textbook in order
to describe the nature of that core. In my preliminary study I examined
only two variables: the age distribution of references, and the total number
of references. Data are presented in table 10. The differences between
the chemistry and physics textbooks, on the one hand, and the sociology
textbooks, on the other, are striking. The overwhelming preponderance
of material discussed in undergraduate chemistry and physics textbooks
was produced prior to 1960. Only 6% of the references in chemistry texts
and 3% of those in physics texts were to work published after 1959. The
exact opposite is true of sociology textbooks. Here I found that fully 75%
of all references were to work published after 1959. Although it might
be argued that chemistry and physics are indeed older disciplines than
sociology, it is clear that any difference in the distribution of the age of
literature in these fields could not possibly account for this dramatic

132

difference in the age of references in textbooks.[12] We should also note the large difference between the chemistry and physics texts, on the one hand, and the sociology texts, on the other, in the total number of references cited. Whereas authors of natural science textbooks generally cite approximately 100 articles or books in their texts, authors of sociology textbooks cite an average of 800 works in their texts.

Although data must clearly be collected for a wider variety of fields and also for graduate textbooks in those fields where they exist, the data in table 10 do lead us to speculate about the differences between the core of knowledge in such fields as physics and chemistry and the core in such fields as sociology. As indicated by what is taught to undergraduate students, the core in fields like chemistry and physics is made up of a small number of theories and exemplars which have been produced by scientists over the course of several hundred years. There is almost universal agreement about which are the important theories and exemplars. In fact, the list of references in the five physics textbooks examined were characterized by substantial overlap, with a majority of the names appearing in all five texts. The same seemed to be true for the two chemistry texts examined. When we look at the sociology texts, we can see that most references are to a wide variety of empirical studies on topics of current interest. Prac-

TABLE 10

AGE AND NUMBER OF REFERENCES IN UNDERGRADUATE TEXTS

	Chemistry* (%)	Physics† (%)	Sociology‡ (%)
Before 1700.	7	16	. . .
1700–99	10	8	. . .
1800–49	12	16	. . .
1850–99	26	17	1
1900–19	12	10	1
1920–39	19	19	4
1940–59	8	11	19
Post 1959.	6	3	75
. Total	100	100	100
Mean N references per text	100	74	800

* Based on W. L. Masterton and E. J. Slowinski, *Chemical Principles* (1977); R. E. Dickenson et al., *Chemical Principles*, 3d ed. (1979).

† Based on G. Shortley and D. Williams, *Elements of Physics*, 3d ed. (1961); W. Thumm and D. B. Tilley, *Physics: A Modern Approach* (1970); M. Alonso and E. J. Fenn, *Physics* (1970); K. Greider, *Invitation to Physics* (1973); F. W. Sears, M. W. Zernansky, and H. D. Young, *University Physics*, 5th ed. (1976).

‡ Based on N. Goodman and G. Marx, *Society Today*, 3d ed. (1978); I. Robertson, *Sociology* (1978).

[12] The great majority of all literature in chemistry and physics has been produced in the post–World War II period because of the very rapid exponential growth of these sciences in the past 30 years.

tically all of these studies were conducted within the past 20 years. The textbooks in the two natural sciences clearly pay little attention to work at the research frontier; those in sociology pay primary attention to such work, probably because in sociology the core of knowledge is very small and the frontier is relatively large.

A qualitative comparison of current textbooks in physics and chemistry with ones used 20 years ago indicated that the material covered 20 years ago and that covered today are essentially the same. This is not true of sociology. Sociology textbooks produced in the early 1950s also relied heavily on recently published work, but the current textbooks cite only a small proportion of the work cited in those used 20 years previously. Thus, although fields like physics and chemistry seem to have a stable core of knowledge on which there is substantial consensus, fields like sociology seem to have a very small core of knowledge with a large research frontier. In short, to the extent to which we can use undergraduate textbooks as indicators of the content of the core of knowledge, it appears that the hypothesized differences among the various sciences do exist in the core.

Consensus and Social Processes

In the last part of this paper, I will present a theoretical explanation of why we find similar levels of consensus at the research frontier in the various scientific fields. Or, to put the matter in a different way, Why is it that in a field like sociology, in which the core is relatively small and weak, we find as much consensus at the research frontier as we find in a field like physics, in which the knowledge at the core is more substantially developed?[13] I have shown that, in the natural sciences, there is probably less consensus at the frontier than has been assumed and that, in the social sciences, there is probably more consensus at the frontier than has been assumed. Kuhn (1970), Polanyi (1958), Lakatos (1970), and Ziman (1968) have argued convincingly that some degree of consensus is a necessary though not a sufficient condition for the accumulation of knowledge in science or in any other type of intellectual activity. The most popular view and the one I shall use as an example is that of Thomas Kuhn.

Kuhn makes a distinction between normal and revolutionary science.

---

[13] To avoid confusion it should be made explicit that I did not find high levels of consensus at the research frontiers of the fields I studied. On the other hand, neither did I find an absence of consensus. If getting an NSF grant is half a result of the "luck of the reviewer draw," it is also half a result of consensus on the merits of the proposal. The important point is not whether the glass is half full or half empty, but that it seems to be filled to the same level in both the natural and the social sciences.

Accumulation of knowledge can occur only during periods of normal science which are characterized by the adherence of the scientific community to a paradigm. It is only when scientists are committed to a paradigm and take it as the starting point for additional research that progress can be made. Without agreement on fundamentals, scientists will not be able to build on the work of others and will spend all their time debating assumptions and first principles. One of the arguments made by Kuhn which was found most surprising and difficult to accept by adherents to the traditionalist positivist view of the development of science was that, during normal science, scientists would ignore empirical results and ideas which could potentially break down consensus on the existing paradigm until a new paradigm has emerged. Because consensus is a necessary condition for advance, the scientific community will reject ideas which will "prematurely" break down the consensus.

As empirical support for this hypothesis, Kuhn and others give many examples of scientists' rejection of negative empirical evidence which seems to contradict the validity of a particular theory. If the experimental results do not fit the predictions made from the theory, the theory is not assumed to be incorrect. Rather, the scientist will assume that there was something wrong in the experimental techniques and that a proper experiment will bring about results which do indeed fit the theoretical expectations. In my own research on the field of superconductivity, I found many examples of this. A content analysis of references to the 1957 Bardeen, Cooper, and Shrieffer theory revealed that experimental physicists' results frequently did not fit the predictions made on the basis of the theory. Yet in not one of these cases did the scientist suggest that the BCS theory might not be correct. Scientists brought up in a positivist tradition might find this type of behavior opprobrious. Kuhn, however, sees consensus on a paradigm as necessary for accumulation and, therefore, sees that the scientific community must act to preserve consensus. He, of course, does not suggest that the preservation of consensus is the *conscious* intent of the community. Instead, it is a latent function of commitment to the paradigm.

A very powerful argument for the necessity of preserving consensus was made by the chemist and philosopher Michael Polanyi (1963). Polanyi relates a story from his own experience. When he was a young chemist, he proposed a theory of adsorption which was contrary to the theories currently accepted. At the time the theory was put forth, it was seen by other chemists as "way out" and was essentially ignored. It turned out 25 years later that Polanyi's theory of adsorption was correct. However, Polanyi himself argues strongly that it was right for the scientific community to reject his idea at the time when it was put forth. Most new and contradictory ideas prove to be of little value. If scientists were too

135

willing to accept every unorthodox theory, method, or technique, the established consensus would be destroyed, and the intellectual structure of science would become chaotic. Scientists would be faced with a multitude of conflicting and unorganized theories and would lack research guidelines and standards. So important is the maintenance of consensus, argues Polanyi, that it is better to reject an occasional idea which turns out to be correct than to be too open to all new ideas at the expense of preserving consensus.

Kuhn argued that consensus was necessary for scientific advance and, assuming that there was little or no consensus in the social sciences, took this deficiency as a cause of their presumably slower rate of accumulation. My data show that, at the research frontier, the social sciences display as much consensus as the natural sciences, and that use of the immediacy effect as a measure of accumulation shows no systematic differences between the natural and social sciences. Therefore, retaining the notion that some level of consensus is a necessary condition or functional requirement for doing scientific work, I suggest that consensus at the research frontier is created and maintained by social processes which tend to be similar in the various scientific fields. Clearly, this theory has all the faults which plague functional theories in general, the primary one being that it is difficult to test. It has been, however, a useful heuristic device for thinking about the problem.

How do social processes actually create consensus, maintain it, and allow it to shift as new ideas emerge? I believe that an answer to this question will be found through an examination of the social processes through which ideas are evaluated in science. How do we know what good science is? In part our evaluations are based on our direct reading and analysis of ideas. But to a large extent our opinions of what good science is and who has done good work are based on judgments made by other people or the evaluation system in our field. To demonstrate this fully would require another paper. There is space here to present only a few examples.

In making evaluations, scientists depend heavily on standards internalized in graduate school. Most scientists have been educated at a relatively small number of prestigious graduate departments where they studied with eminent scientists. To some extent the views of these teachers influence the standards adopted by their students and the subsequent evaluations made by the students.

Data from the studies I have conducted on scientific reward systems indicate that evaluations made by scientists of other scientists are influenced by the rated scientist's departmental rank *independently* of the quality and quantity of his or her publications (Cole and Cole 1973). Thus, if we consider two scientists who have published work of approx-

imately equal quality but one of whom is located in a prestigious department such as Harvard's or Chicago's and the other in a less prestigious state university's, in general the work of the scientist in the prestigious department will be more favorably evaluated. We have internalized the judgment of the evaluation system. If Harvard or Chicago thought that X was good enough to hire, he must be better than Y who works at State University. It is in part because our evaluations are based on judgments made by others that we have the process of accumulative advantage. A favorable evaluation at t_1 will influence the evaluation received at t_2 independently of the quality of role performance in the intervening period.

There are many more examples of how our scientific judgments are influenced by evaluations made by others. We read papers in journals only after they have been evaluated by others. We give people more credit for publications in prestigious journals. We think more highly of people who have received grants, fellowships, awards, memberships in prestigious organizations—all based on the evaluation of others.

Consider the current evaluation system in sociology. I have sent questionnaires to sociologists working in Ph.D.-granting departments and have asked them to name the three sociologists under the age of 40 who they believe have contributed most to the field. Among the most frequently mentioned are several sociologists known for their quantitative mathematical contributions. How does a sociologist doing highly quantitative work become eminent? In this case it is possible to rule out at the beginning the hypothesis that reputation is based on the independent evaluations made by the people who answered the questionnaire. The great majority of the respondents do not have the mathematical background to read mathematical papers, and a substantial majority do not have the background required to assess critically papers using sophisticated quantitative methods such as structural equations or log linear models. Thus, in this case, we can see that reputation is based not on a direct reading of the scientist's work but on the evaluations of the work by authorities in the field. The papers have been published in prestigious journals, their authors appointed to prestigious positions in the field. Most people are willing to accept the judgments made of other people by the evaluation system.

Who are the people who perform the evaluations or occupy the gatekeeper roles? They are generally eminent scientists or "stars" who exercise legitimated intellectual authority. It is generally the stars in a field who determine who will be given appointments in prestigious departments, whose articles will be accepted for publication in prestigious journals, who will receive honorific awards and be admitted to honorific societies.

One of the primary mechanisms through which consensus is maintained is the practice of vesting authority in elites. When a new idea is proposed,

it must be evaluated. Should the new idea become part of the consensus or should it be discarded? In the process of evaluation, some opinions count more than others. Generally, the stars of a particular discipline occupy the main gatekeeping roles. By their acts as gatekeepers and evaluators, they determine what work is considered good and what work unimportant. For the gatekeepers to establish consensus, they must have legitimated authority, that is, the people whose work is being evaluated must accept them as legitimate. Legitimacy is granted by virtue of one's being a star. If the gatekeeper positions are filled by "average" scientists, it will be difficult for the authority exercised to be granted legitimacy. It is only when the scientific community sees those exercising authority as deserving of it that the authority will be accepted.

Because of the need for legitimated leaders, I speculate that the stratification systems of the various sciences are structurally organized in roughly the same way, regardless of their state of cognitive development or the degree to which stars have emerged naturally. A field which does not have "natural" stars will have to create them. The created stars will be those producing the best examples of the most fashionable style of work in the field at the time even though their "absolute" contributions may not be as great as those of the stars who have emerged naturally in other fields or the stars in their own field who preceded them.

REFERENCES

Carter, Grace. 1974. *Peer Review, Citations and Biomedical Research Policy: NIH Grants to Medical School Faculty.* Rand Co. Report 1583-HEW. Santa Monica, Calif.: Rand.
Cole, Jonathan R., and Stephen Cole. 1972. "The Ortega Hypothesis." *Science* 178:368–75.
———. 1973. *Social Stratification in Science.* Chicago: University of Chicago Press.
———. 1980. "Which Researchers Get the Grants." *Nature* 279:575–76.
———. 1981. *Peer Review in the National Science Foundation: Phase II.* Washington, D.C.: National Academy of Sciences.
Cole, Stephen. 1978. "Scientific Reward Systems: A Comparative Analysis." Pp. 167–90 in *Research in the Sociology of Knowledge, Science and Art,* edited by R. A. Jones. Greenwich, Conn.: JAI.
———. 1979. "Age and Scientific Performance." *American Journal of Sociology* 84:958–77.
Cole, Stephen, Jonathan R. Cole, and Lorraine Dietrich. 1978. "Measuring the Cognitive State of Scientific Disciplines." Pp. 209–51 in *Toward a Metric of Science: The Advent of Science Indicators,* edited by Y. Elkana, J. Lederberg, R. K. Merton, A. Thackray, and H. A. Zuckerman. New York: Wiley.
Cole, Stephen, Leonard C. Rubin, and Jonathan R. Cole. 1978. *Peer Review in the National Science Foundation: Phase I.* Washington, D.C.: National Academy of Sciences.
Cole, Stephen, Jonathan R. Cole, and Gary A. Simon. 1981. "Chance and Consensus in Peer Review." *Science* 214:881–86.
Conant, James B. 1950. Forward to *Harvard Case Studies in Experimental Science.* Cambridge, Mass.: Harvard University Press.
Hargens, Lowell L. 1975. *Patterns of Scientific Research: A Comparative Analysis of Research in Three Scientific Fields.* Washington, D.C.: American Sociological Association.

Kuhn, Thomas S. 1970. *The Structure of Scientific Revolutions*. 2d ed. Chicago: University of Chicago Press.

Lakatos, Imre. 1970. "Falsification and the Methodology of Research Programmes." Pp. 91–96 in *Criticism and the Growth of Knowledge,* edited by I. Lakatos and A. Musgrave. Cambridge: Cambridge University Press.

Lodahl, Janice B., and Gerald Gordon. 1972. "The Structure of Scientific Fields and the Functioning of University Graduate Departments." *American Sociological Review* 37:57–72.

MacRae, Duncan. 1969. "Growth and Decay Curves in Scientific Citations." *American Sociological Review* 34:631–35.

Merton, Robert K. 1968. *Social Theory and Social Structure*. New York: Free Press.

Polanyi, Michael. 1958. *Personal Knowledge*. London: Routledge & Kegan Paul.

———. 1963. "The Potential Theory of Adsorption." *Science* 141:1010–13.

Price, Derek J. de Solla. 1970. "Citation Measures of Hard Science, Soft Science, Technology, and Non-Science." Pp. 1–12 in *Communication among Scientists and Engineers,* edited by Carnot E. Nelson and Donald K. Pollack. Lexington, Mass.: Heath.

Rubin, Leonard Charles. 1975. "The Dynamics of Tenure in Two Academic Disciplines." Ph.D. dissertation, State University of New York at Stony Brook.

Small, Henry G. 1974. *Characteristics of Frequently Cited Papers in Chemistry*. Philadelphia: Institute for Scientific Information.

Storer, Norman. 1967. "The Hard Sciences and the Soft: Some Sociological Observations." *Bulletin of the Medical Library Association* 55:75–84.

Ziman, John. 1968. *Public Knowledge*. Cambridge: Cambridge University Press.

Zuckerman, Harriet A., and Robert K. Merton. 1973a. "Age, Aging, and Age Structure in Science." Pp. 497–559 in *The Sociology of Science,* by R. K. Merton. Edited by Norman Storer. Chicago: University of Chicago Press.

———. 1973b. "Institutionalization Patterns of Evaluation in Science." Pp. 460–96 in *The Sociology of Science,* by R. K. Merton. Edited by Norman Storer. Chicago: University of Chicago Press.

Beauty as Status[1]

Murray Webster, Jr.
University of South Carolina

James E. Driskell, Jr.
U.S. Navy Human Factors Laboratory

Physical attractiveness (beauty) affects both cognitions about individuals and their interaction patterns. Our proposed theoretical explanation for these phenomena links attractiveness effects to other cases of status generalization such as those produced by race or sex. Many effects of attractiveness can be explained by viewing it as a status characteristic and applying a theory of status characteristics and expectation states proposed and elaborated by Joseph Berger and others. A test of the proposed explanation shows that (1) attractiveness produces predicted differences in both general and specific expectations; (2) attractiveness effects can be modified in combination with additional status characteristics; and (3) neither of the two above results is affected by sex of stimulus individuals or respondents, a differentiation of this explanation from one that relies on sexual or romantic appeal.

Beauty or physical attractiveness affects social life in both pervasive and profound ways. Folklore and intuition tell us it is fortunate to be beautiful and unfortunate to be ugly, and many recent experimental and naturalistic studies document the respective advantages and disadvantages of these in a wide range of situations. Commonsense and professional explanations of attractiveness effects often rest on strictly individual processes such as romantic or sexual appeal, envy, or desire for equity. In contrast to those, we propose the explanation that attractiveness effects derive usually from the structure of society. Beauty (or its opposite) often functions as a status cue; that is, when it activates patterns of widely shared cultural beliefs it is a status characteristic just as race and sex are, meeting the same defining criteria and having most of the same sorts of effects as those other status characteristics.

After reviewing the range of attractiveness effects, we offer the expla-

nation that these represent cases of status generalization, and we analyze them in terms of a theory developed by Joseph Berger and colleagues (Berger, Fisek, and Zelditch 1977; Humphreys and Berger 1981). We identify links between attractiveness and other status phenomena, present results of an independent test of our analysis, including a strategy for intervention to modify attractiveness effects, and finally offer some conjectures on the nature and future importance of attractiveness in social life.[2]

The most general conclusion from research is that the world must be a more pleasant and satisfying place for attractive people because they possess almost all types of social advantages that can be measured. Dion, Berscheid, and Walster (1972), paraphrasing Sappho, gave the phrase "what is beautiful is good" to these effects. Some of these advantages are documented in the following studies.

More than half a century ago, Perrin (1921) found that attractiveness was an important determinant of popularity, an observation which was significant in Waller's (1937) influential analysis of dating and marriage choices. Not surprisingly, attractive college students have more dates than unattractive students (Huston 1973; Adinolfi 1970). Other studies show that attractiveness has important effects in childhood. Attractive schoolchildren are expected by their teachers to achieve higher school marks than unattractive children, and they actually do so (Clifford and Walster 1973); their misdemeanors in school are judged less serious (Dion 1972); and it is predicted that they will have more successful careers (Dion et al. 1972). Attractive adults are thought to have happier marriages than those who are unattractive (Dion et al. 1972), and that expectation seems to be fulfilled (Kirkpatrick and Cotton 1951; Elder 1969; Berscheid, Walster, and Campbell 1972). Essays alleged to have been written by attractive college students receive higher grades than do the same essays when their supposed authors are homely students (Landy and Sigall 1974). Opinions of attractive adults are more likely to be agreed with (Horai, Naccari, and Fatoullan 1974); attractive adults are perceived as holding attitudes similar to respondents (Cavior and Dokecki 1972); and, in disagreement, attractive adults are more likely to prevail over unattractive ones (Dion and Stein 1978). Attractive adults are perceived as having better mental health (Cash et al. 1977), as being more "attentive" in some interview situations (Kleinke, Staneshi, and Berger 1975), and as being

[2] By "attractiveness" we mean beauty or ugliness. This phenomenon is distinct from *attraction,* which some psychologists have used to mean attitude similarity or liking (e.g., Byrne 1971). Byrne, London, and Reeves (1968) concluded that attitude similarity explains more variance in liking than does attractiveness, though other investigators concluded that attractiveness has greater importance (Stroebe et al. 1971; Berscheid and Walster 1974). We are not concerned primarily with the strength of attractiveness—though it is considerable in most studies—but rather with explaining why it has the effects it does.

more likable, friendly, confident, and sensitive (Miller 1970). Attractive adults are even granted larger "personal space" on the sidewalk than are the unattractive (Dabbs and Stokes 1975).

A few studies have found what the investigators termed "disadvantages of attractiveness." For instance, being too pretty activates envy and so seems to be a liability, for both males and females, in finding a college roommate (Krebs and Adinolfi 1975). Accused swindlers may be seen as more dangerous and be given longer sentences when they are attractive, as compared with those who are unattractive (Sigall and Ostrove 1975); though for most crimes where attractiveness is not an asset in the commission, attractive defendants are treated more leniently (Efran 1974). When rated by unattractive raters, attractive people may be perceived as vain and egotistical, as likely to have extramarital affairs and to request a divorce, and as unsympathetic to the oppressed of the world (Dermer and Thiel 1975). Note, however, that while these correlates may not be admirable, what they have in common with the first set of effects is that attractive people are seen as being better at doing something than unattractive ones. They are seen as more accomplished swindlers or as being more successful in finding illicit lovers.

The relationship between attractiveness and other traits runs in both directions. Thus, college students in judging the attractiveness of others after a group discussion rate as most attractive those whose views coincide with their own along a "radical-conservative" dimension (Walster 1971). Gross and Crofton (1977) showed that attractiveness ratings could be manipulated by changing the social desirability of personality descriptions. Owens and Ford (1978) found the same thing true in ratings of females, though not of males. Felson and Bohrnstedt (1979) concluded that, in a classroom setting, the attractiveness ratings children give each other are influenced by perceived GPA and athletic ability, and not vice versa.[3]

Finally, some recent experiments have documented important effects of attractiveness on interaction. Cash and Kehr (1978) found strong evidence that attractive psychological counselors are seen as more competent, and are more influential and possibly more successful, than unattractive counselors. Reingen et al. (1978) showed that when salespeople were

[3] However, some measurement difficulties prevent the drawing of firm conclusions about the causal model of this study. For instance, athletic ability, which was rated only for boys, varied only from 0 to 4, based on teacher's marks in physical education class, and it may be that the teacher restricted the range of actual grades even more. Felson and Bohrnstedt's main conclusion, that causality can run *from* ability *to* perceived attractiveness, seems sound, but in view of the great amount of other evidence, it seems premature to reject the causal links from attractiveness to many other abilities. Campbell (1979) gives a more elaborate assessment of techniques used in the Felson and Bohrnstedt study.

introduced on the telephone and identified to the person they called only by photos showing them to be attractive or unattractive, they were more successful at closing deals and left a more favorable impression on clients when shown as attractive. Similar findings were reported by Baker and Churchill (1977): attractive advertising models were more effective than plain ones in winning customers.

It would not be difficult to expand the list of studies in which attractiveness appears, but those given above are representative of the literature and illustrate several points which any theory of attractiveness should be able to explain.[4] First, attractiveness produces a wide range of effects—beautiful people have a great many advantages over ugly people—and those effects appear among diverse populations and in a wide range of situations. Second, the attractiveness effects mentioned here all have to do with skills at performing something: attractive people do something better than unattractive ones, whether it be producing a happy marriage, finding partners in adultery, writing essays, or getting people to agree with them. Third, attractiveness affects not only perceived abilities but also interaction, so that attractive people are actually more successful at producing mental health, selling soap, and wielding influence. Finally, the effects are reciprocal, in that attractive people are seen as better at most things and are more successful; and better and more successful people become, by virtue of those traits, more attractive to others.

THEORETICAL VIEWS

Before offering a theoretical explanation for the above findings, we note a surprising lack of such analyses to date. With the notable exception of Walster and her colleagues, no one has conducted a sustained, systematic investigation of attractiveness. Studies which report a single covariate of attractiveness appear, but no investigation of the range or scope of the phenomenon follows them.

The most likely explanation for lack of sustained research on attractiveness is not lack of interest but the lack of an explicit theoretical context for such work. Without the basis of a better developed theoretical understanding, research is noncumulative and fragmented, and any covariates reported have the status of magic. We have no way of predicting whether attractiveness will have effects in a new situation or what those

[4] A comprehensive review of literature to 1974 is provided in Berscheid and Walster (1974), who, along with their students and colleagues, have investigated most of the phenomena reported here. Walster, Berscheid, and Walster (1976) proposed that equity considerations may be responsible for some attractiveness effects in dating choices; i.e., that people seek dates whose attractiveness approximately equals their own: "Intimate relations . . . seem to be influenced in part by equity considerations and in part by fantasy" (p. 36).

effects will be—and no way of understanding either successful or unsuccessful attempts to produce, modify, or eliminate further effects.[5]

The starting point for our analysis is to note that, although many different effects of attractiveness have been documented, most of these can be grouped into effects on *cognitions* regarding attractive and unattractive people and effects on *behaviors* in interaction with those people. As examples of cognitions, we include inferences that attractive people write better essays, have happier marriages, and make better swindlers than unattractive ones. As examples of behaviors, we include the greater likelihood of attractive people's being agreed with by others, exerting more influence, and being granted larger spaces on the sidewalk. The cognitive advantages attractive people enjoy involve attributions of ability: pretty people are thought to be able to do something or "most things" better than ugly people. The behavioral advantages enjoyed by the attractive involve different types of subordination and superordination.

Looked at this way, the advantages enjoyed by attractive people resemble certain advantages men often have in interacting with women or white people have in interacting with black people. That is, the effects of attractiveness seem similar to the effects of the cultural stereotypes that are associated with race and sex. We propose to view attractiveness as a *diffuse status characteristic* like race and sex and to explain its effects through a theory of status generalization proposed by Berger et al. (1977) and developed by Humphreys and Berger (1981). By doing so, we can organize, interrelate, and explain most of the otherwise unrelated findings described above.

PHYSICAL ATTRACTIVENESS AS A DIFFUSE STATUS CHARACTERISTIC

This proposed analysis has two parts; the first is an instantiational claim that physical attractiveness is a diffuse status characteristic in our culture, and the second is a set of theoretical claims about how status characteristics affect cognitions and behavior. Below we present the theory we propose to apply to attractiveness situations, including scope conditions, definitions of key concepts involving diffuse and specific status charac-

---

[5] A possible additional reason for the lack of sustained research may be an emotional aversion to thinking about the advantages beauty confers. We have adages such as "beauty is only skin deep"—i.e., it is unimportant and should not matter in people's lives. However, in describing our research to students and colleagues, we have often been surprised by the anxiety it has aroused. Many researchers in the field write apologetically, as if, in documenting effects of attractiveness, they are uncovering unpleasant truths, and most find themselves so unwilling to use the appropriate words "beautiful" and "ugly" to describe the stimuli of their studies that they substitute awkward circumlocutions such as "more physically attractive" and "those possessing attractiveness to a lesser extent." It is a great deal like refusing to study advantages possessed by rich people because of a strong conviction that such advantages should not exist.

teristics, and five general assumptions.[6] The following is a simplified statement of the theory of status characteristics and expectation states.

Scope conditions.—(1) Any number of actors, including interactants and referent actors; (2) any number of specific and diffuse status characteristics possessed by actors; (3) belief in a task T and a specific ability C^* which is necessary to achieve success at T; (4) motivation to attain success at T; (5) motivation to work collectively at T.

Definition 1.—A characteristic C is a specific status characteristic ▬ (1) the states of C are evaluated differentially in the culture; (2) each state (high or low) of C has a corresponding similarly evaluated expectation state relevant to a particular task.

Definition 2.—A characteristic D is a diffuse status characteristic ▬ (1) the states of D are evaluated differentially in the culture; (2) each state (high or low) of D has a corresponding set of similarly evaluated specific characteristics associated with D; (3) each state of D has a corresponding similarly evaluated general expectation state (GES).

Assumption 1 (salience).—If any D or C status characteristic discriminates individuals (i.e., they possess different evaluated states of that characteristic), such a characteristic will become salient in the task situation.

Assumption 2 (burden of proof).—(1) For any D characteristic which becomes salient in the situation, its associated general expectation state (GES) will be activated and will become relevant to a similarly evaluated state of the specific instrumental characteristic C^* needed for task completion. (2) For any C characteristic which becomes salient in the situation, its relevant task outcome state will be activated and will become relevant to a similarly evaluated state of abstract task ability. That will become relevant to a similarly evaluated outcome state of the group task T.

Assumption 3 (sequencing).—A given structure will be developed through the salience and burden of proof processes for all interactants. If an additional person becomes an interactant, then the structure will be developed further, if necessary, in the same way. For any actor, those parts of his structure completed in relation to a former interactant remain, even if the former interactant leaves the situation.

Assumption 4 (aggregating expectation states).—The relative expectations associated with each actor are calculated as follows. First, all elements of the structure (D and C characteristics and the elements induced by assumption 2) leading to positive expectations are combined to yield a positive value ϵ^+; then all elements leading to negative expectations are combined to yield a negative value e^-; and then aggregated expectations e, for the actor p are given by:

$$e_p = e_p^+ + e_p^-.$$

So p's "expectation advantage" over actor o is $e_p - e_o$.

Assumption 5 (effects of expectations).—Given that any actor p has formed

---

[6] This presentation of the theory is adopted, with some simplifications, from Humphreys and Berger (1981). For full exposition, please refer to that article. For recent illustrations of the generality of the theory, please see the following *AJS* articles: Yuchtmann-Yaar and Semyonov (1979) analyze differential success rates in athletic teams in terms of status generalization based on ethnic differences; Pugh and Wahrman (1983) apply the theory to neutralize one type of sexism among college students.

expectations for self and any other interactant *o, p*'s power and prestige position relative to *o* will be a direct function of *p*'s expectation advantage over *o*.

The scope conditions of the theory require that individuals be oriented toward performance of some task, but in this performance they may be either interactants (such as those granting or withholding deference) or referents (such as those whose supposed essays are being marked). Attractiveness effects which we hope to explain by this theory are limited to those in which people are doing something well or poorly or are deciding how well someone else did something.

Definition 1's "specific status characteristics" are those linked to particular abilities of limited scope: "famous author," "doctor," "class dunce," and "Olympic medalist" are examples. These characteristics are prestigious or invidious (part 1 of the definition) and carry connotations either of some specific skill or of its lack (part 2 of definition). Thus, one expects a doctor to be good at healing and comforting but not necessarily good at repairing automobiles, intercepting a football pass, or logical thinking.

Definition 2's "diffuse status characteristics" include race, sex, age, and, we propose, physical attractiveness. These are prestigious or invidious (part 1), carry connotations of possessing several different additional characteristics (part 2), and also carry connotations for being good at "most tasks" (part 3). Sex is a diffuse status characteristic in our culture because men are widely believed to be better than women at tasks requiring strength, mechanical skill, logic, athletic ability, distance judgment, etc. That *etc.* is important because diffuse status characteristics carry connotations of characteristics and abilities without any explicit limit.

To construct a detailed explanation of the physical attractiveness phenomenon, we begin by claiming that attractiveness meets the definition of a diffuse status characteristic in our culture. Then, by assumption 1, whenever attractiveness differentiates individuals, it will become salient—noticed, treated as a significant fact about their interaction. Assumption 2, which is in many ways the most intriguing part of the theory, says that, whenever a status characteristic is salient, people will treat it *as if* it gave useful clues to the ability to perform tasks. That is, whether or not there is any "logical" reason to believe that the status characteristic is relevant to successful task completion, people will treat it as if it were relevant. In other words, the burden of proof is placed upon demonstrating that status is *not* relevant to ability, instead of the other way around.

Assumption 3 states that structures are completed through salience and burden of proof processes. (Complicated interactions, in which people enter and leave the situation, do not apply here.) Assumption 4 gives a

function for combining different status information in order to arrive at aggregated or overall expectations for each actor. While the exact mathematical form of the function is not at issue here, the claim that all status information is used in status generalization—none is ignored or eliminated through cognitive processing—is important, especially in cases where individuals know of other status characteristics (such as age, sex, occupation, or various abilities) besides attractiveness. Assumption 5 is the basic expectation assumption that translates expectations into observable behavior. It says that all behavioral features of subordination and superordination—such as agreement, influence, and being given a large space on the sidewalk—vary directly with the aggregate expectations associated with each particular actor.

Thus, the theory posits a process with the following form:

$$D \longrightarrow GES \longrightarrow \text{specific expectations} \longrightarrow \text{interaction inequality.}$$

Differentiating status characteristics (D) become activated and lead through burden of proof to an induced "general expectation state" or GES for each actor. The GES in turn leads to formation of task-specific expectation states, ideas about doing well or poorly at some task; these expectations then determine observable interaction advantages for all actors in the situation.

To use the theory to explain attractiveness phenomena, we must first show that attractiveness can be interpreted reasonably as meeting the three parts of definition 2 and then that the effect being explained is either a specific expectation (cognition) formed through assumptions 1, 2, 3, and 4 or a behavior determined through assumptions 1, 2, 3, 4, and 5.

If attractiveness is a diffuse status characteristic, it must possess at least two states which are evaluated differentially. All research shows that this criterion is fulfilled: there are both attractive and unattractive people, and it is preferable, better, advantageous, and more desirable to be pretty than to be ugly.

Second, there are specific, evaluated traits associated with states of the characteristic, traits that cause attractive people to be seen as performing certain specific tasks better. Attractive students receive higher expectations for performance in school (Clifford and Walster 1973); attractive counselors are more "expert" (Cash and Kehr 1978); attractive adults are thought to have more successful marriages (Dion et al. 1972)—these are specific expectations for a wide range of different tasks.

Third, there is a very general expectation state such that attractive people are seen as more skilled at *most* tasks. The total picture of evidence could be said to show that attractiveness meets this criterion. However, there is more direct evidence in a few studies that shows that, when

respondents are given a large number of descriptive adjectives, they perceive attractive people as more desirable (Miller 1970) and able to do better at "most things in life" (Dion et al. 1972).[7]

AN EMPIRICAL TEST

It is reasonable to ask of any proposed theoretical explanation not only whether it can provide a post hoc explanation for what is known but whether it can also yield predictions for additional situations within its scope. To test the usefulness of our analysis, we now turn to devising and conducting such an independent test.

To do this, we select photographs of attractive and unattractive people, present those photos to respondents, and attempt to measure both specific expectations (part 2 of the definition of D) and general expectations (part 3). In addition to supporting the instantiational claim (that attractiveness is a status characteristic), we should be able to use our analysis to produce the expectation effects predicted by the theory.

Therefore, this empirical test includes three main tasks: (1) to manipulate effectively attractive and unattractive stimuli; (2) to gauge the effects of attractiveness in producing task-specific and general expectations, according to the definition of a diffuse status characteristic; and (3) to test the usefulness of the theory in modifying predicted attractiveness effects.

For research purposes, a reasonable answer to what is attractive is an empirical one. While most people have difficulty describing what they consider attractive, and while there is always some variability in judges' ratings (Cross and Cross 1971), still, most studies find a fairly high degree of consensus about what is pleasing (Berscheid and Walster 1974), and there is some evidence that, within a given culture, ratings are unaffected by age and SES of the judges (Iliffe [1960], cited in Berscheid and Walster [1974]). Beauty may be like art: we may not know much about it, but we know—and generally agree on—what we like.[8]

---

[7] We do not claim that *all* the effects of attractiveness are explainable as cases of status generalization; only that most of them, including some very important ones, are. Status generalization produces beliefs about abilities of people—such as the ability to write an excellent essay or to carry out a successful swindle—and most of the covariates of attractiveness can be seen to operate through the intervening variable of such ability conceptions. But some effects, such as the demonstration by Goldberg, Gottesdiener, and Abramson (1975) that ugly women are perceived as more likely than pretty women to support feminist movements, are not readily explicable by the approach we propose here.

[8] For operational realization of attractiveness, the literature provides only incomplete guidelines. Feldman ([1971] cited in Berscheid and Walster [1974]) found that height was a considerable advantage for males, while Berscheid and Walster (1974) claim that Dorothy Parker was right about the negative effect of women's wearing glasses. Anthropologists report wide variation in cross-cultural standards of attractiveness, and several scholars find that social factors such as wealth are a greater component of overall desirability for men than for women (Ford and Beach 1951; Miller and Rivenbark 1970; Krebs and Adinolfi

To select photographs that differ reliably in attractiveness, we used a modification of the NORC method of measuring occupational prestige. In the first stage of the process, we examined approximately 800 college yearbook photographs. Because respondents were to be college students themselves, we selected yearbook pictures of graduating seniors as an appropriate place to find examples of prevailing standards of attractiveness. To protect the privacy of people photographed, we restricted the yearbooks studied to campuses several hundred miles away from the respondents'.

The 800 photographs were narrowed to 161 by the two investigators, and then to 59 (27 males, 32 females) by other members of the research staff. The 59 photographs were judged by 25 paid volunteers who were college students from the same population to be used in the eventual data collection. These students arranged photographs into nine piles so that each pile contained photos of approximately equal attractiveness. After sorting the photographs initially, each rater was asked to examine the contents of each pile and to move any photographs which seemed to belong to another pile. From that sorting, photos of two males and two females from either end of the scale were chosen for the eight stimulus photos.

Pictures chosen were those that produced the highest discrimination in scores and the lowest variances among raters. The mean ratings for the attractive males and females, respectively, were 7.68 and 8.06; for the unattractive males and females, 1.58 and 1.10. Variances of ratings ranged from 0.07 (unattractive female) to 2.19 (attractive male).[9]

The second part of our task is gauging the effect of attractiveness in producing expectations. For that, we use a questionnaire developed in the course of previous work (Webster and Driskell 1978) designed to be sensitive to both parts 2 and 3 of the definition of a diffuse status characteristic. In the previous study, questionnaire scores correlated 0.62 ($P < .001$) with other status information, including race and ability (Driskell and Webster, forthcoming).

Pairs of same-sex photos were reproduced on questionnaires containing the expectation measures. Each questionnaire thus contained one photo

---

1975). Seeing bilateral asymmetry—someone whose body's right and left halves are not mirror images of each other—may be distressing to most people (Heider 1981). Our strategy here is to start with strong differentiation on the attractiveness dimension; we return to consideration of exactly what constitutes beauty in the Discussion section of this paper.

[9] To continue the comparability to NORC ratings, we could transform the pile numbers given here to ratings as follows: $(X - 1) \times 12.5 = $ rating, where X is the average number of the pile into which a given photo is placed. That gives attractiveness ratings from 93.75 to 1.25 for the individual photos, or an average of 85.87 for the four attractive ones and 4.25 for the four unattractive ones (see Siegel [1971] for description of the NORC rating procedure).

of a person who was rated highly attractive and one who was rated highly unattractive. The instructions and questions are reproduced below.

1. How intelligent do you perceive Person A to be, as compared to Person B?

A is much more intelligent	A is somewhat more intelligent	They are exactly equal	B is somewhat more intelligent	B is much more intelligent
(1)	(2)	(3)	(4)	(5)

(Same scoring for all answers)

2. Compared to Person B, how well do you expect Person A to do in situations in general?
3. In terms of things that you think count in this world, how do you compare Person A to Person B?
4. How capable do you think Person A is compared to Person B at most tasks?
5. How do you compare Person A with Person B concerning reading ability?
6. How do you compare Person A with Person B at abstract abilities?
7. How would you compare Person A's grade point ratio (GPR) with Person B's?
8. Both Persons A and B took the FAA exam for a private plane pilot's flying license. How well do you think they probably did on this exam?
9. In your opinion, how do these two people compare in overall physical attractiveness?

Questions 1–7 were intended to measure general expectations (the *GES* of the theory) or performance at tasks in general. Question 8 is designed to measure expectations for performance at one specific task (piloting a plane)—a task suitable for either males or females, for which attractiveness is not a relevant characteristic, and one with which the respondents are presumed not to have personal involvement. Question 9 was included as a check on the manipulation of attractiveness.

By assumption 2, beauty will affect expectations so long as it is not definitely known to be unrelated to the task at hand. Thus a beautiful person will be expected to be likely to win a beauty contest (a relevant task) or to be a better pilot (a task for which beauty has no direct relevance). Cases of the second sort (burden of proof) have greater substantive interest—although they do not differ significantly in theoretical terms from relevant cases—so that is the type we attempted to create in our design. Further, effects of attractiveness should also be altered in predictable ways by techniques similar to those used to modify effects of the characteristic race. More specifically, if attractiveness is indeed a diffuse status characteristic, we know that it should combine with any other status characteristics in a situation (assumptions 1 and 4), so that the perfor-

150

mance expectations formed are a composite result of all available status information (Cohen and Roper 1972; Webster and Driskell 1978). Thus, we should be able to lessen the expectation disadvantage of an ugly person by simultaneously introducing other high-status cues. Similarly, the higher expectations induced by a beautiful stimulus person should be tempered by the introduction of additional low-status cues.

The sequence of administration of the study was as follows. A member of the research staff addressed a class of respondents and told them we were interested in studying social perception. They would be given a questionnaire designed to examine how impressions of others are formed on the basis of limited visual and social information. Class members were also informed they were under no compulsion to participate in the study and could either simply not take a questionnaire or leave it blank.

In order to strengthen the task orientation of respondents (a scope condition of the theory of status generalization), our researcher told them that the people depicted in the questionnaires had previously taken part in a longitudinal study at several colleges and had completed a follow-up study several years after graduation. They were also told that the study in which they were participating would enable us to compare the accuracy of each respondent's impressions of Person A and Person B with the actual performances of these individuals.

Each questionnaire presented a pair of pictures for comparison; the information presented on each questionnaire varied in amount and type according to each of six conditions. In condition 1, respondents rated a highly attractive stimulus photo. In condition 2, respondents rated the photo of the low-attractive person. In condition 3, the photo of the high-attractive person was rated, but he or she was also described as having graduated from a low-prestige school (a small, regional state college) and being employed in a low-status occupation (clerk at Woolworth's). In condition 4, respondents rated the photo of the low-attractive person, who was described as having graduated from a high-prestige (Ivy League) school and being employed in a high-status occupation (vice-president at Chase Manhattan Bank). Conditions 5 and 6 are similar to conditions 3 and 4 in that respondents were given the same descriptions, but in conditions 5 and 6 there was no attractiveness manipulation; thus the respondents saw no pictures. In condition 5, respondents rated the high-status person; in condition 6, they rated the low-status person. Table 1 shows the information given for all six conditions.

The requested ratings were comparative: thus, a single questionnaire contained both one attractive and one unattractive person's photo for conditions 1 and 2. Questionnaires were collated to distribute them among respondents, with right or left placement of photos rotated systematically throughout.

151

Persons A and B on each individual questionnaire were of the same sex and were contrasted in status information (attractiveness of person in photo in conditions 1, 2, 3, and 4; status of occupation and college in conditions 3, 4, 5, and 6). Respondents were asked to examine the photos and descriptions, and then to give their impressions on the items shown in the questionnaire on *GES* and expectations presented earlier. In addition, the respondents were asked to report their own sex, major, age (four categories), and race (black, white, other).

Respondents were told that since the questionnaires differed each person should concentrate on the one he or she received and ignore what other people in the room were doing. Reports from the administrators indicated that the respondents undertook the rating task with a high degree of interest and made a serious attempt to rate the people accurately. It was not possible to calculate how many members of the classes chose not to complete or not to turn in their questionnaires, but our impression was that virtually everyone in the classes completed and turned in a questionnaire.

PREDICTIONS

To test the instantiational claim that attractiveness is, indeed, a diffuse status characteristic, we predict that the expectations, as measured here,

TABLE 1

STATUS INFORMATION AND INSTRUCTIONS TO RESPONDENTS

| | STATUS ELEMENTS | | |
CONDITION	Attractive-ness	Educational Status	Occupational Status
1	High	...	...
2	Low	...	...
3	High	Low	Low
4	Low	High	High
5	...	High	High
6	...	Low	Low

NOTE.—Attractiveness manipulation: High attractiveness/low attractiveness. Educational status: Graduation from Ivy League school/small regional college. Occupational status: Employed as corporate officer in finance/variety-store clerk.

Instructions: "The following questions deal with your impressions of the two persons pictured above. We are interested in how people use visual and social clues in forming impressions, so please study the photographs and read the descriptions before answering the questions. While some of the questions may seem hypothetical in nature and hard to answer on the basis of limited information, they are important; so please be as honest and frank as possible in answering them. You may refer back to this page at any time. Circle the number on the scale that best represents your feelings."

will be higher in condition 1 than in condition 2. To show further that the expectations produced by attractiveness can be modified by additional status information (alma mater and occupation), we also predict higher expectations in condition 4 (unattractive, high-status college and occupation) than in condition 2 (unattractive person's picture only). Similarly, the expectation advantage of the photo of the attractive person in condition 1 can be reduced by introducing low educational and occupational status in condition 3. Finally, on the basis of much previous work, we predict that the education and occupation characteristics by themselves will cause expectations to be ordered $5 > 6$ in those conditions.

RESULTS

After removing incomplete questionnaires and those that were from respondents over the age of 30 (fewer than 10 of each), we had 415 usable questionnaires. Removing the questionnaires of 56 respondents who identified themselves as black left 359 for analysis.[10] Thus, 180 usable questionnaires deal with conditions 1 and 2 (photos only); 118 deal with conditions 3 and 4 (photos plus status information); and 61 deal with conditions 5 and 6 (status information only). In the following analyses, data were transformed so that the "focal individual" is placed in the position of Person B; for example, when we report data for the unattractive person's photo (condition 2), they are presented as if that person occupied position B, even though he was presented half the time in position A. This has the effect of identifying higher numbers with higher expectations instead of the reverse. The expectation measure is simply the sum of points checked on each item in the scale.

First, we examine properties of the attractive and unattractive persons' photos. To justify collapsing pairs of stimulus people into single categories—for example, "attractive female," "unattractive male"—a Mann-Whitney U-test is applied to responses on the item 9 attractiveness check for each set of same-sex pairs. That shows no significant differences (P's $> .10$) except for the two unattractive females ($P \simeq .04$), a finding which

---

[10] The issue of race has not been raised in previous research on attractiveness, though obviously standards of what is considered attractive vary. We decided to restrict the photos and respondents to white since that group makes up more than 80% of the student body at the university where the research was conducted. The number of black respondents excluded—56, distributed across the six conditions of the study—is too small for detailed separate analysis. However, in general, it appears that the following two conclusions are supported. First, black respondents perceived smaller differences between the attractive and unattractive stimulus photos of their questionnaires than white respondents did; and second, given that difference, status generalization effects were similar for black and white respondents. We are indebted to Professor A. Wade Smith of Arizona State University for being the first to point out to us that what blacks think is attractive is often not the same as what whites find attractive, with particular reference to our stimulus photographs.

may be due to the fact that those photos had received the two lowest variances of the eight in the original sortings. Between-condition differences are substantial; the mean attractiveness rating in condition 1 is 4.62; in condition 2 it is 1.25 ($P < .01$); in conditions 3 and 4 the ratings are 4.40 and 1.31, respectively ($P < .01$). An analysis of variance shows a significant main effect for attractiveness ($P < .001$), with no significant effects of sex of respondent, sex of stimulus person, or interaction. We conclude that attractiveness was manipulated as intended and that it is justifiable to speak of "attractiveness" as a characteristic in the following analyses without specifying sex of person in photo or sex of respondent.

Finding no significant effects by sex of person in photo or sex of respondent is important, for it is consistent with the claim that attractiveness effects are due to the fact it is a status characteristic. A competing argument, that attractiveness effects depend on sexual or romantic appeal, for instance, would lead to a prediction that sex of respondent and of person in photograph would interact in the majority of cases. But the status interpretation does not require that respondents find the attractive persons' photos appealing—just that they see them as cues to high status.[11]

A varimax rotation factor analysis of the eight-item scale for measuring general expectations is presented in table 2. This shows two factors. The first includes situations in general, things that count in this world, most tasks, abstract ability, and the specific task of piloting a plane. Factor 2 includes reading, intelligence, and grades. We will call the first factor "general ability" and emphasize it in the analyses, since it seems to reflect more accurately the theoretical definition of a diffuse status characteristic. Factor 2, "scholastic ability," will be discussed afterward.

TABLE 2

FACTOR LOADINGS OF WITHIN-CELLS
CORRELATION MATRIX

	VARIMAX ROTATED FACTORS	
VARIABLE	Factor 1	Factor 2
Intell	.374	(.810)
Sitgen.	(.754)	.134
Count.	(.766)	−.020
Tasks	(.651)	.365
Read	.131	(.792)
Abstr	(.413)	.283
GPR.	.221	(.857)
SPC	(.397)	.146
ATT.	(.583)	−.621

[11] A study published since this research was completed (Anderson and Bem 1981) also reported no difference in response to attractiveness by sex of respondent.

Figure 1 shows the mean scores by condition for the factor 1 items, with scores ranging from 5 (lowest expectations) to 25 (highest). All predicted orderings of conditions are observed. In conditions 1 and 2, attractive people are assigned higher expectations than unattractive ones ($Z = 7.72$, $P < .001$). In conditions 5 and 6, expectations are formed through status generalization from the educational and occupational status characteristics and are higher for the graduates of an Ivy League college with a high-status job than for graduates of a regional college with a low-status job ($Z = 5.27$, $P < .001$).

In some ways, the most important conditions are 3 and 4, where we attempt to alter the high expectations attributed to attractive people and the low expectations attributed to unattractive people. In condition 3, a highly attractive person is described as a graduate of a low-status regional college who is holding a low-prestige job. The expectation advantage the highly attractive person of condition 1 enjoyed is now reduced in condition 3 (17.3 in condition 1 vs. 14.3 in condition 3; $Z = 7.02$, $P < .001$). In condition 4, an unattractive person is described as coming from a high-status college and holding a high-prestige job. Here, the lower expectations assigned to the unattractive person are raised (12.8 in condition 2 vs. 16.5 in condition 4; $Z = 5.73$, $P < .001$).[12]

Comparing the expectation difference between conditions 1 and 2 with the difference between conditions 5 and 6 shows greater differentiation in the latter two conditions. That is what we would expect from the theory of status generalization: in conditions 1 and 2 only a single status
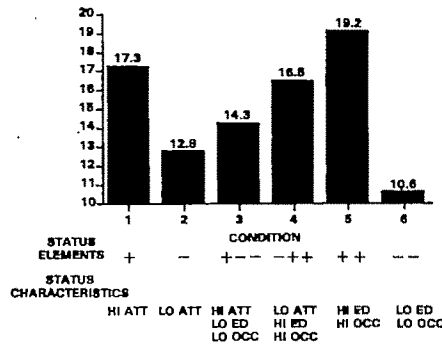
FIG. 1.—Expectations by condition (factor 1)

[12] There is some justification in looking at the results of item no. 8, piloting a plane, separately, for the definition of a diffuse status characteristic requires that it carry both general expectations (such as those in items 2, 3, 4, and 5) and specific task expectations (such as item 8). In condition 1, the mean score on item 8 was 2.13; in condition 2, it was 1.61. So pretty people are thought to make better pilots than ugly people—just as the theory of status generalization would predict from the knowledge that attractiveness is a status characteristic.

characteristic, attractiveness, differentiates Person A from Person B, while in conditions 5 and 6, there are two, college and occupation. Since all available status information is combined to form aggregate expectations, expectation differences produced by two status characteristics should be greater than differences produced by one (see Humphreys and Berger [1981], Theorem 2, for rigorous derivation of this prediction).

A particularly interesting comparison may be made between conditions 3 and 6, and 4 and 5, respectively. In both pairs of conditions, the social status information is the same: low-prestige college and job in both conditions 3 and 6, and high-prestige college and job in both conditions 4 and 5. But the expectations attributed to the attractive stimulus person in condition 3 (14.3) are somewhat higher than the expectations attributed to the person in condition 6 (10.6). Similarly, expectations attributed to the unattractive stimulus person in condition 4 (16.5) are lower than expectations attributed to the person in condition 5 (19.2). So the attractiveness characteristic has expectation effects despite the presence of the more reasonable bases of occupation and college for judging ability. An ugly graduate of a high-prestige college who holds a high-prestige job is assumed to be somewhat less capable than another graduate of the same college holding the same job whose picture is not shown. Also, a pretty graduate of a low-prestige college holding a low-prestige job is accorded higher expectations than someone else whose appearance is unknown from the same school holding the same occupation. In other words, even when college of graduation and occupation are known, people still also take into account the attractiveness of someone when deciding how well he does at, say, piloting a plane, or "most tasks."[13]

Finally, we return to the issue of bidirectional causality between attractiveness and abilities noted earlier. Explaining bidirectionality of effects is not difficult from the theory of status generalization. From this viewpoint, people "infer" ability levels from any known status characteristics, so any personal characteristic that fits the definition of a status characteristic can produce high performance expectations for any given task. For the more common causal direction, if someone is highly at-

[13] Here we also find support for theorem 3 of Humphreys and Berger. This theorem states that, given equally relevant discriminating characteristics, the greater the degree of inconsistency of these characteristics, the less the degree of differentiation. In other words, the greater inconsistency of characteristics in a system, the less inequality—actors possessing either all consistently high (white, male, attractive) or consistently low (black, female, unattractive) characteristics would be differentiated to a greater degree than if they each possessed inconsistent characteristics, some high and some low. Consider our conditions 5 and 6, in which actors are discriminated on consistent characteristics, and conditions 3 and 4, in which actors are discriminated on inconsistently evaluated characteristics. We see that there is a greater degree of inconsistency for conditions 3 and 4, and, as this theorem claims, we find less differentiation between these conditions than we find between conditions 5 and 6.

tractive, status generalization will produce high performance expectations for him at almost any specific task. This is the direction most research emphasizes. To explain the reverse direction, we can readily suppose that most ability measures actually represent status characteristics, either specific ones (such as "excellent reader" or "famous author") or diffuse (such as "successful person" or "welfare recipient"). Thus those status characteristics should produce generalization to specific tasks, and it may be that people think that being attractive is something that people can succeed or fail at. The theory of status generalization essentially describes a process by which people "infer" (sometimes not consciously) the possession of other, similarly valued characteristics from characteristics that are known. Thus we would expect that people who possess the high state of some status characteristic would, by that fact alone, come to be seen as also more likely to possess the high state of attractiveness, and vice versa. In support of this analysis and consistent with the earlier studies, we do find generalization from certain status characteristics to perceived attractiveness. This shows up in responses to question 9, the perceived attractiveness check, for conditions 5 and 6 where no picture was presented. In condition 5, the high-status description, Person B received an average attractiveness rating of 3.52. In condition 6, the low-status description, Person B received an average attractiveness rating of 2.68 (P < .05).[14]

To return briefly to factor 2, intelligence or scholastic ability, results show that this factor is correlated negatively with perceived attractiveness of photos as measured by item 9 ($-.621$), perhaps suggesting a "beautiful but dumb" stereotype among our college student respondents. Some of these volunteered comments such as "Person B [attractive] probably has a lot of things to do, so I guess Person A [unattractive] gets better grades." Whether that finding reflects only a social stereotype of college students or whether it would reappear in other settings is unknown.

Yet factor 2 was virtually unaffected by actual attractiveness of the photos. For example, the mean rating on factor 2 for condition 1 (attractive

[14] This explanation still does not fit the one study which found that ability perception affected perceived attractiveness, not vice versa (Felson and Bohrnstedt 1979). Except for the possible methodological consideration in n. 3 above, we cannot explain conclusively why those investigators did not detect status generalization from attractiveness to ability. It may be that children's ideas of each others' abilities are determined by the teacher's marks, in which case status generalization could not modify those judgments much. Or it might be that the range of perceived attractiveness among children studied was not great enough for status generalization to operate in most cases. See also Campbell (1979) for the argument that, because of collinearity between socioeconomic status and school marks, it may be that socioeconomic status and not marks was the real factor affecting perceived attractiveness. (That is, that children find middle-class children more attractive than working-class ones.) If that is true, then the Felson-Bohrnstedt negative result might be caused by the fact that the appropriate dependent variable—perceived socioeconomic status—was not measured.

photo) was 5.4; for condition 2 (unattractive) it was 6.9 ($P > .05$). Conditions 3 and 4 differed on factor 2, as did conditions 5 and 6, but the differences appear to be produced entirely by the social status information, education and occupation (condition 3 = 3.2, condition 6 = 2.5, condition 4 = 9.2, condition 5 = 9.6). Neither of those differences is significant.

Those findings weaken the "beautiful but dumb" interpretation in its application to actual attractiveness, especially because the item 9 attractiveness check from which it comes appeared after all the expectation-measuring questions. What we found, then, was that respondents said they thought that highly intelligent people were less attractive than people of lower intelligence. But when presented with a pair of people who actually differed greatly in attractiveness, respondents estimated very little difference in their intelligence.

DISCUSSION

The major purpose of this investigation was to offer a theoretical context for much of the existing research on effects of physical attractiveness. We proposed that many—not all—effects of physical attractiveness can be explained by recognizing that attractiveness is a status characteristic, and, as such, produces ideas of task competencies. In other words, the theory of status generalization may be applied to explain many attractiveness phenomena.

Results of this study show that attractiveness meets the theoretical definition of a diffuse status characteristic in our culture. That is, (1) the states are evaluated differentially—it is considered better to be beautiful than to be ugly; (2) people who possess the high state of attractiveness are also expected to possess the high state of other specific characteristics—in this case, they are considered more competent to pilot a plane; and (3) people who possess the high state of attractiveness are also expected to possess the high state of general, unlimited characteristics—in this case, they are seen as better at "situations in general, things that count in this world, most tasks, and abstract ability."

Results also document effects on perceived skills that have virtually no intrinsic relation to beauty or ugliness. Thus the burden of proof predictions of the theory are supported in a stringent test. The fact that attractiveness effects function equally in the case of male and female subjects and that there is no interaction of sex of rater and sex of subject may help to differentiate predictions of our status generalization theory from alternative theories that rely on implied romantic or sexual attraction. Finally, the bidirectional relationships between the attractiveness characteristic and other characteristics which have been reported by some

recent investigators become interpretable through the theory of status generalization.

What is the practical use of conceptualizing attractiveness as a status characteristic? One answer is in the results from conditions 3 and 4 of this study, in which the attractiveness characteristic was paired with an inconsistent set of additional status characteristics. From the point of the theory, attractiveness, occupations, and education all are status characteristics—though of course they differ greatly in their concrete realizations. Seeing attractiveness and, for example, education as different manifestations of the same underlying construct, however, tells us how to expect them to function in situations. Because status characteristics combine, expectations produced through status generalization reflect an aggregate view of all the status characteristics possessed by each person. So we can intervene to modify the effects of attractiveness, using any of several intervention techniques developed to modify effects of other status characteristics, particularly race (Webster and Driskell 1978; Cohen, Katz, and Lohman 1976; Cohen and Sharan 1980) and sex (Ridgeway 1982; Lockheed 1980).

Note that not all interventions are aimed at reducing the interaction advantages possessed by highly attractive people. For example, it is common practice in sales to try to maximize the effects of status generalization by hiring "sharp-looking salespeople." Attractiveness would be predicted to be especially important where a customer relies upon a salesman's expertise, as is often the case in sales of complicated office equipment such as computers, or when independent insurance agents and stockbrokers advise clients which policy or investment to buy. Job seekers should similarly note that it is distinctly in their interest to activate as many high-status elements as possible, including attractiveness, in job interviews.[15]

Given that the goal is to overcome status generalization, the theory suggests several strategies, one of which was demonstrated in this research. Intervention here focused on introducing high-status information for the unattractive subjects. We know that expectations are formed, not on the basis of a single status characteristic, but rather from the entire set of statuses an individual possesses. If someone possesses the low state of attractiveness (or sex, race, or age), it becomes especially important to make salient some compensating high statuses. Thus it may be desirable

---

[15] Graduate students facing job interviews are not always receptive to suggestions that they conform to prevailing definitions of attractive dress and grooming. The story is told that Talcott Parsons's advice to one student was, "Since you intend to revolutionize sociology, and perhaps society, let me offer one suggestion. Wear a tie. You can sneak up on them that way."

to "add on" high-status elements such as clothing or visible diplomas, certificates, etc.

An issue that deserves further investigation is the extent to which attractiveness effects produce a "self-fulfilling prophecy" for the attractive or unattractive people themselves. Snyder, Tanke, and Berscheid (1977) and Reingen et al. (1978) both reported that, when unseen people are thought to be attractive by those with whom they are talking, the supposedly attractive individuals actually behave in more likable, socially desirable ways. The suggestion is that attractive people receive cues from others which actually make them behave better; the reactions that pretty people get may, in fact, make them learn to be more adept socially. Thus they get a double advantage: from status generalization and from learning social skills.

While there is a great popular literature on just what is attractive, scientific evidence for such claims is scant. In asking what beauty is, with special reference to its effects as a status characteristic, we are struck by two features, *conventionality* and *status*.

Conventionality appears to be the most important component of beauty, both in our study and in others. The eight pictures selected to represent attractive and unattractive subjects in our study could reasonably be said to represent the normal range of college students on that dimension, since they were chosen from over 800 original yearbook pictures. Yet there is no obvious characteristic that either the attractive or the unattractive ones have in common. The two attractive men look neat and clean (both have styled hair that is over their ears and is parted in the middle, have no facial hair, and are wearing white shirts and ties); open and friendly (one is smiling); and otherwise quite average. The two attractive women are also neat and clean (both wear long hair parted in the middle and look "wholesome" rather than "glamorous"), open and friendly (one is smiling widely, one with her lips closed), and otherwise quite average. The unattractive men both wear glasses, their hair is worn over their ears but has not been trimmed recently, smile and look friendly, and wear white shirts with ties. The unattractive women do not wear glasses, have hair only slightly less groomed than that of the attractive women, smile broadly, and show more facial fat than the attractive women. The overall impression from these pictures is that very minor and superficial changes in hairstyle, weight, and, perhaps, clothing would be sufficient to make the unattractive people more attractive, or the reverse. In other words, the pretty ones probably looked the way their parents wanted them to, while the ugly ones had ignored one or more pieces of simple advice.

Ford and Beach (1951) noted that standards of sexual attractiveness, which have some overlap with physical attractiveness, vary widely in different cultures. One thing considered unattractive in all cultures is

dressing in clothing worn by the opposite sex. Indeed, one of the unattractive women in this study was wearing a white shirt with a tie. We cropped the tie out of the picture before using it, but it is possible her face still looked "masculine" to our respondents. Other observers have noted how much European and American standards of beauty have changed over time; for example, an amount of body fat that was considered voluptuous in the 19th century is considered unfortunate now, and our contemporary standard of thinness would have been looked on as evidence of tuberculosis as recently as 40 years ago.

But if conventionality is a key to attractiveness, getting reliable advice on what the prevailing norms are is not always easy. Most women learn eventually that men do not want them to wear as much makeup as fashion models or to be quite that thin. High school boys have to learn that copying the look of teenage rock singers is not a sure route to being considered attractive. Opinion polls show regularly that many well-known men (and a few women) are considered highly attractive by national samples of respondents, though from their photographs alone it is difficult to say just why. So while there seem to be some influential norms of attractive appearance, developing a clear statement of what those norms are at any particular time is difficult.[16]

The second component of attractiveness, shown in this research and others, is status on some other dimensions. It has been remarked jokingly that having a million dollars can make someone much more attractive, though the suggestion there is that what really happens is that people overlook ugliness in those who have money. But if we are correct about the status generalization functioning of attractiveness, what happens is somewhat different from that. Someone with the high state of some status dimension really will be perceived as more attractive than, say, an identical twin who lacks such high status. Thus "the good are beautiful" would be more accurately stated as "the high-status people are beautiful."

Another way in which status may influence attractiveness is through setting contemporary standards of beauty. Clothing fashions and hairstyles are designed for the rich, and it is almost always upper-socioeconomic-status individuals whose appearances others try to copy.[17] Something

[16] Note also that our pictures, like virtually all of those used in previous research, focus on head and shoulders. Conceptions of beauty usually include impressions of the entire body, though whether full body photos would show a greater range of perceived attractiveness is unknown. Attractiveness might not function as a status characteristic at extreme ranges; it might induce pity at one extreme or jealousy at the other, for instance. These and other issues may benefit from further inquiry.

[17] Veblen ([1899] 1973) observed that a prime requisite of clothing that invites invidious comparison is that it be impractical for any kind of work. More recent observers (Wolfe 1970; Brooks 1981) have identified the fashion for blue jeans as a sort of "reverse snobbism," *nostalgie de la boue*. However, it should be noted that the most prestigious (and thus, we

of this sort may operate in classrooms where students prefer the appearance of middle-class children to that of working-class ones. Developing a strong sense of personal "style" or being made a cult figure like Humphrey Bogart also helps an otherwise plain person become attractive.[18]

The reason we see the components of attractiveness—what is considered beautiful—as less important than its operation—through status generalization—is that the former change greatly in a short period of time. The theoretical process of status generalization should be invariant through time and culture. That is, whatever meets the definition of a diffuse status characteristic in a particular time and place will have predictable effects; and so long as attractiveness is a status characteristic, whatever the prevailing standards of beauty are, "beautiful" people will have certain strong advantages as a result. Those advantages are similar to the advantages conferred by other status characteristics such as sex and race, and their operation can be influenced by principles understood from study of those other statuses.

The importance of physical attractiveness in people's lives is only beginning to be recognized. Berscheid and Walster (1972) quote Aristotle that "beauty is a greater recommendation than any letter of introduction," a claim that now has theoretical and empirical support. Beauty or ugliness is one of the most accessible features of a person and acts as readily available status information in most encounters. While discrimination because of race or sex may decline because of changes in laws and educational practice, discrimination based on attractiveness evidences no comparable movement.

An important difference between attractiveness and other diffuse status characteristics such as race or sex, however, is that it is to some extent an achieved characteristic. Minor change can be effected through clothing or cosmetics, while reconstructive or plastic surgery is available for major alterations. The malleability of attractiveness is an additional reason to expect an increasing amount of attention directed toward this characteristic. Certainly the growth of private gyms—where the explicit appeal may be to health, but the implicit appeal is to beauty—of tanning clinics, and the sales figures for books of advice on makeup and dressing for success show an awareness on the part of the consumers of these services

would predict, the most "attractive") kinds of denim clothing are "designer" jeans, which are not often worn for hard work.

[18] Beauty pageants seem to show quite clearly the operation of conventionality and status. Our colleague Lala Steelman reports that her students usually make quite accurate predictions about the height, weight, hair color, and hobbies of successful Miss America contestants—predictions which are dependent on Americans' conventional ideas about beauty. Since, usually, about one-third of the entrants in such international contests as Miss Universe have dark skin, yet virtually none of the finalists do, dark skin is a status characteristic. We are grateful to Professor Steelman for these notes.

that attractiveness is important. Because status generalization defines interactional situations, it may be that attractiveness will assume increasing significance as other characteristics such as race and sex fall into disuse.

REFERENCES

Adinolfi, A. 1970. "Characteristics of Highly Accepted, Highly Rejected, and Relatively Unknown University Freshmen." *Journal of Counseling Psychology* 17:456–64.

Anderson, S. M., and S. L. Bem. 1981. "Sex Typing and Androgyny in Dyadic Interaction: Individual Differences in Responsiveness to Physical Attractiveness." *Journal of Personality and Social Psychology* 41:74–86.

Baker, M. J., and G. A. Churchill. 1977. "The Impact of Physically Attractive Models on Advertising Evaluations." *Journal of Marketing Research* 14:538–55.

Berger, J., M. H. Fisek, and M. Zelditch, Jr. 1977. *Status Characteristics and Social Interaction: An Expectation States Approach.* New York: Elsevier.

Berscheid, E., and E. Walster. 1974. "Physical Attractiveness." Pp. 157–215 in *Advances in Experimental Social Psychology*, vol. 7, edited by L. Berkowitz. New York: Academic Press.

Berscheid, E., E. Walster, and R. Campbell. 1972. "Grow Old along with Me." Mimeographed. Cited in Berscheid and Walster 1974.

Brooks, J. 1981. *Showing Off in America.* Boston: Atlantic, Little, Brown.

Byrne, D. 1971. *The Attraction Paradigm.* New York: Academic Press.

Byrne, D., O. London, and K. Reeves. 1968. "The Effects of Physical Attractiveness, Sex, and Attitude Similarity on Interpersonal Attraction." *Journal of Personality* 36:259–71.

Campbell, R. 1979. "The Relationship between Children's Perceptions of Ability and Perceptions of Physical Attractiveness: Comment on Felson and Bohrnstedt's 'Are the Good Beautiful or the Beautiful Good?'" *Social Psychology Quarterly* 42:393–98.

Cash, T. F., and J. Kehr. 1978. "Influence of Nonprofessional Counselors' Physical Attractiveness and Sex on Perceptions of Counselor Behavior." *Journal of Counseling Psychology* 25:336–42.

Cash, T. F., J. Kehr, J. Polyson, and V. Freeman. 1977. "Role of Physical Attractiveness in Peer Attribution of Psychological Disturbance." *Journal of Consulting and Clinical Psychology* 45:987–93.

Cavior, N., and P. R. Dokecki. 1972. "Physical Attractiveness, Perceived Attitude Similarity, and Academic Achievement as Contributors to Interpersonal Attraction (Popularity) among Fifth and Eleventh Grade Boys and Girls." Mimeographed. Cited in Berscheid and Walster 1974.

Clifford, M. M., and E. Walster. 1973. "The Effect of Physical Attractiveness on Teacher Expectation." *Sociology of Education* 46:248–58.

Cohen, E. G., M. L. Katz, and M. R. Lohman. 1976. "Center for Interracial Cooperation: A Field Experiment." *Sociology of Education* 49:47–58.

Cohen, E. G., and S. S. Roper. 1972. "Modification of Interracial Interaction Disability: An Application of Status Characteric Theory."*American Sociological Review* 37:643–55.

Cohen, E. G., and S. Sharan. 1980. "Modifying Status Relations in Israeli Youth." *Journal of Cross-cultural Psychology* 11:364–84.

Cross, J. F., and J. Cross. 1971. "Age, Sex, Race and the Perception of Facial Beauty." *Developmental Psychology* 5:433–39.

Dabbs, M., Jr., and N. A. Stokes. 1975. "Beauty Is Power: The Use of Space on the Sidewalk." *Sociometry* 38:551–57.

Dermer, M., and D. L. Thiel. 1975. "When Beauty May Fail." *Journal of Personality and Social Psychology* 31:1168–76.

Dion, K. K. 1972. "Physical Attractiveness and Evaluations of Children's Transgressions." *Journal of Personality and Social Psychology* 24:207–13.

Dion, K. K., E. Berscheid, and E. Walster. 1972. "What Is Beautiful Is Good." *Journal of Personality and Social Psychology* 24:285–90.

Dion, K. K., and S. Stein. 1978. "Physical Attractiveness and Interpersonal Influence." *Journal of Experimental Social Psychology* 14:97–108.

Driskell, J. E., Jr., and M. Webster, Jr. Forthcoming. "Status Generalization: New Data," in *Status, Attribution, and Rewards,* edited by J. Berger and M. Zelditch, Jr. San Francisco: Jossey-Bass.

Efran, M. G. 1974. "The Effect of Physical Appearance on the Judgment of Guilt, Interpersonal Attraction, and Severity of Recommended Punishment in a Simulated Jury Task." *Journal of Research in Personality* 8:45–54.

Elder, G. H., Jr. 1969. "Appearance and Education in Marriage Mobility." *American Sociological Review* 34:519–33.

Feldman, S. D. 1971. "The Presentation of Shortness in Everyday Life—Height and Heightism in American Society: Toward a Sociology of Stature." Paper presented at the annual meeting of the American Sociological Association, August, Denver.

Felson, R. B., and G. W. Bohrnstedt. 1979. "Are the Good Beautiful or the Beautiful Good? The Relationship between Children's Perceptions of Ability and Perceptions of Physical Attractiveness." *Social Psychology Quarterly* 42:386–92.

Ford, C. S., and F. A. Beach. 1951. *Patterns of Sexual Behavior.* New York: Harper.

Goldberg, P. A., M. Gottesdiener, and P. R. Abramson. 1975. "Another Put-Down of Women?: Perceived Attractiveness as a Function of Support for the Feminist Movement." *Journal of Personality and Social Psychology* 32:113–15.

Gross, A. E., and C. Crofton. 1977. "What Is Good Is Beautiful." *Sociometry* 40:85–90.

Heider, K. G. 1981. "Good Form." Mimeographed. Department of Anthropology, University of South Carolina.

Horai, J., N. Naccari, and E. Fatoullan. 1974. "The Effects of Expertise and Physical Attractiveness upon Opinion Agreement and Liking." *Sociometry* 37:601–6.

Humphreys, P., and J. Berger. 1981. "Theoretical Consequences of the Status Characteristics Formulation." *American Journal of Sociology* 86:953–83.

Huston, T. L. 1973. "Ambiguity of Acceptance, Social Desirability, and Dating Choice." *Journal of Experimental Social Psychology* 9:32–42.

Iliffe, A. H. 1960. "A Study of Preferences in Feminine Beauty." *British Journal of Psychology* 51:267–73.

Kirkpatrick, C., and J. Cotton. 1951. "Physical Attractiveness, Age, and Marital Adjustment." *American Sociological Review* 16:81–86.

Kleinke, C. L., R. A. Staneshi, and D. E. Berger. 1975. "Evaluation of an Interviewer as a Function of Interviewer Gaze, Reinforcement of Subject Gaze, and Interviewer Attractiveness." *Journal of Personality and Social Psychology* 31:115–22.

Krebs, D., and A. A. Adinolfi. 1975. "Physical Attractiveness, Social Relations, and Personality Style." *Journal of Personality and Social Psychology* 31:245–53.

Landy, D., and H. Sigall. 1974. "Beauty Is Talent—Task Evaluation as a Function of the Performer's Physical Attractiveness." *Journal of Personality and Social Psychology* 29:299–304.

Lockheed, M. E. 1980. "Expectation States and Status in Mixed-Sex Groups: A Review." Paper presented at the annual meetings of the American Sociological Association, August, New York.

Miller, A. G. 1970. "Role of Physical Attractiveness in Impression Formation." *Psychonomic Science* 19:241–43.

Miller, H. L., and W. H. Rivenbark III. 1970. "Sexual Differences in Physical Attractiveness as a Determinant of Heterosexual Likings." *Psychological Reports* 27:701–2.

Owens, G., and J. G. Ford. 1978. "Further Consideration of the 'What Is Good Is Beautiful' Finding." *Social Psychology* 41:73–75.

Perrin, R. A. C. 1921. "Physical Attractiveness and Repulsiveness." *Journal of Experimental Psychology* 4:203–17.

Pugh, M., and R. Wahrman. 1983. "Neutralizing Sexism in Mixed-Sex Groups." *American Journal of Sociology* 88:746–62.

Reingen, P. H., J. B. Kernan, L. Gresham, C. Narashmhan, and I. Ronkainen. 1978. "Cognitive and Behavioral Consequences of the Physical Attractiveness Stereotype in Personal Selling." Mimeographed. School of Business, University of South Carolina.

Ridgeway, C. 1982. "Status in Groups: An Application of Status Characteristics Theory." *American Sociological Review* 47:76–88.

Siegel, P. M. 1975. "Prestige in the American Occupational Structure." Ph.D. dissertation, University of Chicago.

Sigall, H., and N. Ostrove. 1975. "Beautiful but Dangerous: Effects of Offender Attractiveness and Nature of the Crime on Juridic Judgment." *Journal of Personality and Social Psychology* 31:410–14.

Snyder, M., E. D. Tanke, and E. Berscheid. 1977. "Social Perception and Interpersonal Behavior: On the Self-fulfilling Nature of Stereotypes." *Journal of Personality and Social Psychology* 35:656–66.

Stroebe, W., C. A. Insko, V. D. Thompson, and B. D. Layton. 1971. "Effects of Physical Attractiveness, Attitude Similarity, and Sex on Various Aspects of Interpersonal Attraction." *Journal of Personality and Social Psychology* 18:79–91.

Veblen, T. (1399) 1973. *The Theory of the Leisure Class*. Boston: Houghton Mifflin.

Waller, W. 1937. "The Rating and Dating Complex." *American Sociological Review* 2:727–37.

Walster, E. 1971. "Did You Ever See a Beautiful Conservative?" Mimeographed. Cited in Berscheid and Walster 1974.

Walster, E., E. Berscheid, and G. W. Walster. 1976. "New Directions in Equity Research." Pp. 1–42 in *Advances in Experimental Social Psychology,* vol. 9, edited by L. Berkowitz and E. Walster. New York: Academic Press.

Webster, M., Jr., and J. E. Driskell, Jr. 1978. "Status Generalization: A Review and Some New Data." *American Sociological Review* 43:220–36.

Wolfe, T. 1970. *Radical Chic and Mau-Mauing the Flak Catchers*. New York: Farrar, Straus & Giroux.

Yuchtman-Yaar, E., and M. Semyonov. 1979. "Ethnic Inequality in Israeli Schools and Sports: An Expectation States Approach." *American Journal of Sociology* 85:576–90.

Max Weber, "Rational Capitalism," and Renaissance Italy: A Critique of Cohen[1]

R. J. Holton
Flinders University of South Australia

Contrary to Cohen, Max Weber did not believe that the initial development of modern rational capitalism occurred only after the Reformation. In the case of Renaissance Italy, however, recent empirical studies confirm Weber's view that the scale of modern rational capitalist development was still limited and that economic traditionalism prevailed. Although some technical advances were evident (e.g., in accounting), their significance has been exaggerated. This is due in large measure to the use of simplistic evolutionary models of social change which (in Weber's terms) conflate "technical rationality" with "rational economic action."

Jere Cohen's (1980) article in the *AJS,* "Rational Capitalism in Renaissance Italy," represents an important new challenge to the much debated connection that Max Weber drew between Protestantism and the development of the spirit of modern capitalism. Cohen's intention is not to claim that capitalism in some unspecified general sense is older than Protestantism (the strategy of some previous critics [e.g., Tawney 1926, p. 84]), but to show "that modern capitalism, *in Weber's sense,* existed in pre-Reformation Catholic countries" (Cohen 1980, p. 1342). This challenge rests not so much on the conceptual analysis and reworking of such notions as "capitalism" or "rationality" as on the deployment of a vast mass of data drawn from historical studies of pre-Reformation Italy.

The aims of this paper are twofold. First, I argue that Cohen underestimates the extent to which Weber was prepared to accept the existence of elements of modern rational capitalism prior to the Reformation. In this sense, the energy Cohen devotes to rebutting the allegedly Weberian proposition "that rational capitalism developed initially and primarily under Protestantism" (Cohen 1980, p. 1340) is misdirected. It serves merely to perpetuate the myth of Weber's supposedly monocausal "Protestant ethic" explanation of the development of modern capitalism. Attention is thereby deflected from Weber's far broader concern with analyzing the contrasts and continuities *within* Occidental history, between the ancient, medieval, and modern epochs.

Second, I argue that Cohen's attempt to portray the considerable scale of rational capitalism within Renaissance Italy is grossly exaggerated. While his argument is supported by an impressive collection of citations which open up some important issues in historical sociology, it is nonetheless deficient in four important respects. First, Cohen presents an extremely loose and incomplete account of Weber's notion of modern capitalism. Second, he all but ignores Weber's comments on the pre-Reformation economic history of European capitalism. Third, he misconceives the implications of Weber's ideal-type method in the analysis of historical development. Fourth and last, his handling of the empirical data is, in certain key instances, seriously inaccurate and lacking in contextual sensitivity.

MAX WEBER AND MODERN CAPITALISM

On the conceptual level, Cohen rightly stresses Weber's distinction between the older (adventurer and fiscal-monopoly) capitalism and modern "rational" capitalism. Weber characterizes the latter, in Cohen's terms, by its rationality in business organization, accounting methods, and technology and by the presence of free markets and a calculable legal system. Cohen does not, however, elaborate Weber's discussion of rationality and modern capitalism any further, leaving considerable uncertainty about the criteria whereby Weber applied the concept of rationality to modern capitalism.

The varied and overlapping senses in which the category of rationality has been used by Weber and others makes this a notoriously elusive and ambiguous concept to define and operationalize in empirical research. Nonetheless, Weber attempted some important distinctions in his discussion of rationality, economic action, and capitalism.

In one sense, rationality, viewed in terms of the existence of a means-end relationship, is seen by Weber as a formal property of all types of social action. In another looser, commonplace sense, rationality approximates to Weber's notion of "instrumental rationality" centered on social action involving "rational consideration of alternative means to the end, of the relations of the end to the secondary consequences, and finally of the relative importance of different possible ends" (Weber 1978, 1:26). In this sense, rationality is often associated with action oriented to find the best technical means to reach a given end.

Weber, however, also made a further important distinction between "rational technique" and "rational economic action" (Weber 1978, 1:65). Thus "the technique of an action refers to the means employed as opposed to the meaning or end to which the action is, in the last analysis oriented" (Weber 1978, 1:65). This distinction enables Weber to come to terms with

167

the fact that significant degrees of "rationalization" (i.e., technical rationalization) could be compatible with widely contrasting orientations from instrumental rationality to traditional rationality "ingrained by habituation" (Weber 1978, 1:69–70). Rational economic action does not reside in certain formal techniques alone. Instead, it depends on substantive changes in the orientation of economic action itself. The foremost of these is the shift from the age-old capitalistic activity involving "an impulse to acquisition, the pursuit of gain, of money," etc. (Weber 1958, p. 16) to modern capitalism. This involves the pursuit of profit and forever renewed profit by means of *continuous* rational capitalist enterprise as an end in itself.

Because Cohen has almost completely neglected these considerations, his criteria for the empirical presence of modern rational capitalism "in Weber's sense" are not only vague but also seem to be somewhat wider than Weber's. They appear to include, for instance, types of economic activity which, while utilizing certain rational technical forms (e.g., double-entry bookkeeping), nonetheless rarely sustain the renewed search for profit on a continuous basis. At any rate Cohen certainly leaves the door open for inclusion of Renaissance business activities in which the "rational pursuit" of profit was neither a guiding principle of life nor regarded as an end in itself. It was seen instead as one of a number of means of supporting some other life principle such as the status aspirations of the traditional medieval household.

Cohen's treatment of Weber is also vulnerable on the historiographical level. For while he concedes that Weber "knew that some capitalistic practices predated the Reformation" (Cohen 1980, p. 1340, citing Weber 1958, p. 91), no real effort is devoted to establishing further what Weber had in mind. Having apparently assumed that he must be referring to adventurer and fiscal-monopoly capitalism, Cohen goes on to attribute to Weber the belief "that the calculating spirit was applied to acquisitive activities only in recent times" (Cohen 1980, p. 1342). From this it follows—at least for Cohen—that Weber "was unaware of the *extent* to which rational capitalism had developed in pre-Reformation Italy" (Cohen 1980, p. 1340). Hence this account of Weber's "lack" of historical sense ignores key texts such as *Economy and Society* and Weber's doctoral dissertation "On the History of Medieval Trading Companies." It also glosses over highly pertinent comments made by Weber in the texts Cohen himself cites, namely, *The Protestant Ethic* and the *General Economic History*.

In an important footnote within the *Protestant Ethic*, Weber rebutted the proposition "that capitalism as an economic system is a creation of the Reformation" with the rather cryptic counterproposition that "certain forms of capitalist business organization are known to be considerably

older than the Reformation" (Weber 1958, p. 91). However, other texts—notably, the *General Economic History* and Weber's early work on medieval trade—indicate in more detail that such "forms" relate not only to adventurer and fiscal-monopoly capitalism but also to the new types of mercantile institution developing in precisely those Renaissance cities, such as Genoa and Florence, with which Cohen is concerned.

Weber drew particular attention to the significance of the widespread use of credit rather than cash in the complex long-term trading ventures of the Italian merchants. This in turn encouraged the development of new legal forms of business organization to guarantee credit. These took the form of various types of rational partnership in which the extent of individual liabilities was defined and limited and the apportioning of risks was clarified. Weber also argued that such organizational developments, coupled with the incorporation of outside capital in hitherto wholly family concerns, stimulated a tendency away from the traditional household organization of business. This was, however, only a tendency, since newly emerging forms of business organization coexisted with the more traditional household organization of commerce for a long while. In this respect the separation of business from the household was far from complete. This indicated to Weber that Renaissance business had not yet succeeded in creating an economy dominated by modern capitalism, though it had anticipated certain features of the new social system (Weber 1924, 1961).

Weber went on to make a clear distinction between economies (such as that of Renaissance Italy) which were organized partly capitalistically, containing "somewhat capitalistic establishments" (Weber 1961, p. 207), and the modern post-18th-century epoch, in which the capitalist organization of daily wants had become predominant. In the former case, as with the Renaissance merchant houses that subsequently disintegrated as commercial units, it is possible to conceive the removal of such enterprises from the economic life of the time without introducing any overwhelming change (Weber 1961, p. 207). In the latter case, however, it is inconceivable to imagine the removal of capitalist enterprises without the collapse of the entire system. For Weber this highlights an important historical shift from the pre-Reformation world, where the "calculating spirit" existed merely as a means of economy in a context where other general rules of conduct prevailed, to a situation in which this spirit itself had become "a principle of general conduct."

In this context the "Protestant Ethic argument" was not meant to be used, as Cohen supposes, to explain the *initial* development of various forms of the "calculating spirit" or rational economic behavior. Many of these predated the Reformation and in Weber's view required consideration of other developments in Occidental history, such as Roman law (Weber 1961, pp. 250–52) or the development of the autonomous late

medieval city (Weber 1978, 2:1322–39), in order to explain their emergence. The function of the Protestant ethic was, rather, to provide one strand in the exploration of how, after the Renaissance, a more fundamental social change occurred whereby economic rationality and calculation emerged as a dominant end in itself.

Cohen suggests that the reason for Weber's failure to appreciate the extent of pre-Reformation capitalism, and the consequent continuity between pre- and post-Reformation capitalism, has to do with deficiencies in the ideal-type method (Cohen 1980, p. 1350). The objection that Weber tended to hypostatize ideal-type concepts, thereby destroying "the organic unity of concrete historical phenomena and the historical process" (Parsons 1949, 2:607), is a familiar one. It is doubtful, however, whether Weber's reluctance to interpret pre-Reformation mercantile activity as anything more than "anticipatory" of modern capitalism can be explained in terms of the general difficulty that Parsons has raised. Weber's own work on the medieval period (Weber 1924, 1961), with its careful consideration of complex processes of historical change, should certainly dispel any belief that he was unwilling to admit the existence of widespread deviations from the "pure" ideal-type of modern capitalism. In this respect Weber appears to have been consistent in following his methodological prescription concerning the status of ideal-types as conceptual abstractions helpful in framing questions but not as empirical generalizations or hypotheses in their own right.

What seems more germane to the explanation of Weber's position is the critical stance he adopted toward evolutionary theories of social change (Roth 1976, pp. 309–10). Weber's neo-Kantian epistemology, according to which the course of "history" was inherently meaningless, led him to resist evolutionary accounts of the development of capitalism, which, as in the case of late-19th-century Marxism, assumed some kind of objective teleological unfolding of history toward a generally predetermined end. At the same time this did not prevent him from accepting the notion of developmental history (*Entwicklungsgeschichte*). This alternative to evolutionism permitted Weber to depict the logic of modern capitalism, and more generally of rationalization, while at the same time leaving room to analyze its complex historical development in terms of contingent features of Occidental society rather than historical necessity.

Although Cohen does not utilize an explicitly evolutionary perspective, he does come close to invoking a Whig theory of capitalist emergence against Weber. This presents past historical epochs such as the Renaissance merely as stages leading to the present. Thus a "long . . . historical process of development" is posited as culminating in "the finished system" (Cohen 1980, p. 1350). For Weber, in contrast, pre-Reformation capitalism cannot be considered an early stage in the "ever onward and upward"

development of modern capitalism. Pre-Reformation capitalism contains no inner dynamic leading toward the modern system. It is not Weber's ideal-type method but this antievolutionary standpoint which explains Weber's reservations about the developmental significance of the Renaissance economy.

RATIONAL CAPITALISM IN RENAISSANCE ITALY

Cohen's discussion of business practice in 14th-, 15th-, and 16th-century Italy is based on a valuable synthesis of a large body of monographic research. Nonetheless, when we move from an initial inspection to a closer scrutiny of his data base, a number of serious problems of poorly grounded inference and textual distortion emerge. The following empirical critique concentrates on one major section of Cohen's paper, namely, "The Spirit of Rational Calculation" (Cohen 1980, pp. 1345–47). This seems appropriate since the German historical sociologists of capitalism, such as Weber and Sombart ([1915] 1967, 1919), placed such emphasis on the "spirit" of modern capitalism and its relation to the emergence of new institutions.

Within this section there is no explicit definition of what does and does not constitute rational calculation; hence it is difficult to establish just how modern capitalism might differ from traditionalism. Instead Cohen begins by presenting some very general statements on the calculated ordered attitude of Italian merchants. These are taken to be self-evident examples of a modern rational spirit as opposed to some older (yet unspecified) form of social action.

There are grounds for skepticism about a number of Cohen's procedures for the interpretation of evidence cited here. In the first place, he supports interesting arguments, for example, that Italian merchants "customarily did not waste time" (Cohen 1980, p. 1345), with a single very general and rather vague citation. Although the authority cited on this occasion (Sapori 1969, p. 75) clearly has a detailed knowledge of business records, the actual passage on which Cohen bases his claim simply records Sapori's impression that merchants, in contrast with some other social groups, were "not used to wasting time." This may well be true. However, on the evidence presented, it scarcely supports by itself Cohen's implied suggestion that a *new* rational attitude to time had been created in Renaissance Italy as part of the "spirit of rational calculation."

Such claims have been made for late medieval Western Europe by Le Goff (1960). He contrasts "church time" based on a theological view of the cosmos with "merchant time" based on exact measurement and cast in a rationalized secular framework. However, this interpretation draws more on French than on Italian evidence and is in any case subject to a

number of powerful criticisms. In the first place, the empirical basis of Le Goff's claim is rather dubious. It depends in part on the erection in major cities of public clocks, which he emphasizes were used to indicate the hours of commerce and employment. While Le Goff sees the public clock as an instrument of economic domination, part of the weaponry of emergent capitalist class struggle (Le Goff 1960, p. 425), Cipolla (1967) presents a more complex picture. Public clocks were as much the product of civic rivalry and aesthetic pride on the part of local elites (Cipolla 1967, pp. 42–43) as utilitarian instruments. Acceptance of their use may well have paved the way for the later post-Reformation diffusion of domestic clocks when technological change finally cheapened the cost of production. At the same time Le Goff's suggestion that the medieval public clock presents an antecedent to 20th-century Taylorist work discipline seems farfetched. It appears, for example, that medieval clocks were extremely inaccurate, "lost or gained much time in a day (if and when they worked)," a phenomenon Cipolla relates to "low" contemporary requirements for precision. Minute hands were thus a rarity. Meanwhile the traditional sundial continued to meet most needs, including the resetting of inaccurate public clocks.

There is obviously room for legitimate debate about changing conceptions of time and their relationship to capitalist development. Nonetheless, even when Cohen's conceptually abbreviated and empirically sparse reference is supplemented with the additional data presented above, there remains little hard evidence that Italian Renaissance merchants had gone very far toward the thoroughgoing rationalized attitude to time implied in Benjamin Franklin's 18th-century adages "time is money" and "never put off until tomorrow what you can do today." It is far more likely that the shift from a traditional conception of time oriented around the seasons and tasks associated with them, to the modern precise measurement of time, most notably as a means to labor productivity, really occurred around the 17th and 18th centuries with the emergence of wage labor organized outside the household and within the factory (Thompson 1974).

A second line of criticism of Cohen's initial treatment of the spirit of rational calculation concerns his indifference toward establishing the intellectual contexts within which the authorities he cites are operating. Of the major authorities cited most frequently (i.e., those acquainted with primary data), only a very few, such as Raymond de Roover, expressly intended their work as a rebuttal of Max Weber's supposed thesis concerning the Reformation and capitalism (de Roover 1963, p. 7). In the main, the work of Italian historians such as Sapori or Luzzatto was directed against Sombart's general view of the medieval merchant as a small-scale artisan, lacking foresight, planning, and education (Sapori 1953, pp. 53–58; Luzzatto 1953, pp. 50–52). Within this context, rebuttal

172

of Sombart does not appear to require the same criteria to be met concerning the nature of economic activity as are necessary for the rebuttal of Weber. In other words, it is not necessary to prove a continuous rational orientation to profit seeking as a guiding principle of conduct. All that is required is to show that merchants operated on a large scale, planned their activities, were educated and literate, and had some organized conception of business accounting. Hence it seems that a good deal of the historiography of Renaissance life that Cohen utilized to buttress his case is actually concerned more with establishing the presence of profit-seeking mercantile capitalism than with demonstrating the existence of modern rational capitalism in *Weber's sense.*

Having established in this rather loose sense that medieval merchants and businessmen sought profits through market calculation, Cohen goes on to argue that production and marketing methods were themselves rational and, by implication, modern. This judgment is based on his view that "standardized products were developed to satisfy an unknown, abstract circle of purchasers" (p. 1346).

Of the five sources cited in support of this view, the first (Ehrenberg [1928] 1963, p. 55) appears to argue the reverse of what Cohen assumes. The passage cited concerns the organization of exchange and the timing of the shift from irregular exchange organized through fairs to regular exchange organized through daily commerce. It is quite true that Ehrenberg refers to a situation whereby commodities must be so far standardized that they do not have to be seen at the time of transaction! The sentence is, however, only hypothetical, dealing with prerequisites necessary for the shift to daily commerce rather than the situation prevailing during the Renaissance. In practice, with the exception of the bill of exchange, Ehrenberg saw the irregular fair as still the characteristic form of exchange for the wholesale trade of the Middle Ages. The point he is making is that standardization had not so far developed, a situation contrasting with "modern times."

Another authority (de Roover 1963, p. 144) also lends Cohen very little support. De Roover, like Ehrenberg, draws a contrast between medieval and modern trading practice. Medieval merchants, for example, did not usually have orders for goods personally placed; instead, commodities were generally sent by consignment. However, de Roover does not describe this practice as involving "an unknown abstract circle of purchasers" within the businessman's calculations. Rather, he refers to medieval merchants as "adventurers in that they sent their goods . . . without knowing whether or not they would sell" (1963, p. 143). This contrasts with the modern (rational) situation where "the buyer usually seeks out the seller and goods are not shipped without being sold."

The development of "modern" accounting practices such as double-

entry bookkeeping seems at first sight to be a more promising support for Cohen's conception of Renaissance capitalism. It is perhaps ironic that Weber, against whom Cohen is arguing, also testified to the pioneering Italian contribution to "genuine bookkeeping" based on computation in arabic figures, developed in the late medieval period (Weber 1961, pp. 170–74). Nonetheless, since Cohen's case turns on the question of scale as well as qualitative advance, it is important to clarify the soundness of his quantitative estimations about Renaissance bookkeeping as inferred from the massive set of historical authorities cited on medieval accounting practice.

In quantitative terms, Cohen argues that double-entry bookkeeping (i.e., the recording of each transaction on both the debit [liability] and credit [asset] side of the account), "was in general use throughout Italy in the 15th century . . . despite some firms that were exceptions" (Cohen 1980, pp. 1346–47). Given the fragmentary survival of Renaissance account books (de Roover 1956, passim), it is, of course, difficult to be as confident as Cohen is of the general character of existing practice. Nonetheless, even allowing for this, closer examination of the citations indicates Cohen's claim for the general use of double entry to be rather dubious.

Only two of Cohen's eight citations on the use of double entry deal in any substantive way with the quantitative impact of double entry. De Roover (1956) in particular provides a comprehensive survey of virtually all extant account books of the Italian Renaissance. Nearly 30 enterprises are mentioned for the period 1296–1477, yet of these, only seven are adjudged to be definitely in double entry. When looked at over time, however, the picture does show a trend toward double entry with six of the 12 enterprises from the post-1390 period being adjudged to be in double entry, compared with only one out of 14 from the pre-1390 period.

Clearly the proportion of those in double entry, not the absolute numbers, is more relevant in the interpretation of these data. Even so, it is doubtful whether one can speak of double entry being in general use even in the 15th century. Certainly on de Roover's evidence, double entry would have been practiced by only half the sample. Moreover, the surviving records, on which his sample is based, tend disproportionately to favor larger concerns, which it may be hypothesized were more likely than smaller ones to use double entry. This view is consistent with the opinion of Datini, a leading Florentine merchant, who, while using double entry himself, complained of the accounting methods of the smaller merchants of Prato, his town near Florence. Thus the Pratesi kept their books in their heads "like the carriers who reckon up their accounts 20 times along the road. Four out of six of them have neither book nor ink-well, and those who have ink have no pen" (Datini, quoted in Origo 1957, pp. 108–9).

Since Cohen is concerned with establishing the extent rather than simply the presence of double-entry bookkeeping, it seems reasonable to infer the significant and increasing use of double entry over time, especially among the larger firms. At the same time businesses not using double entry, namely, half of those examined by de Roover for the period 1390–1470, can scarcely be regarded as exceptional.

Cohen's assessment of a second aspect of pre-Reformation accounting, use of the trial balance, rests on rather less secure empirical foundations than his exaggerated generalizations about double-entry bookkeeping. There are two aspects to this question: the extent of trial-balance usage among various enterprises and the regularity of usage.

As far as scale of usage is concerned, Cohen cites three authorities to support the proposition that the trial balance was "fully known and in general use at the end of the 15th century" (Cohen 1980, p. 1347). Yet these record only three or four definite cases of trial-balance usage. While the practice no doubt went further than this, by no stretch of the imagination does Cohen's evidence prove its adoption on a widespread scale.

Similar exaggerations are evident in Cohen's discussion of the regular use of trial balances. What is at stake here is neither the drawing of balance at the completion of one ledger to effect transference to another ledger, nor closing balances drawn at either the death of a businessman or his withdrawal from a partnership. Rather, it is the practice of drawing a trial balance during the normal course of business to check the equality of debits and credits, and possibly for wider purposes such as managerial control. Cohen's impressive list of 14 page citations from six sources, however, yields only three positive cases of the regular use of trial balance. Moreover, this has to be set against considerable negative evidence of infrequent usage.

Most of Cohen's claims for the considerable extent of both double-entry bookkeeping and regular trial balances within Renaissance Italy seem exaggerated. Nonetheless, since these new technical forms were present on a limited basis, it remains to establish in what sense their use may be seen as rational and modern.

There are two important aspects to this question. In the first place, there is no guarantee that users of double entry or regular trial balances exhibited a more "rational technique" in Weber's sense than those of their contemporaries who adopted other methods such as single entry. De Roover in particular stresses the accounting difficulties faced by manufacturers in the use of two currencies, gold and silver (the first to pay suppliers, the second to pay labor), between which the exchange relationship fluctuated considerably (de Roover 1956, pp. 138–50). It is also probable that smaller businesses with a less complex scale of operations than the much-quoted larger merchants found it less appropriate to use complex ac-

175

counting procedures. In the case of trial balances, it also seems that regular assessment of assets and liabilities was somewhat inappropriate for merchants engaged in long-term international trading ventures lasting several years.

This argument suggests that rational technique in Weber's sense cannot simply be regarded as embodied in certain techniques (e.g., double entry). Instead, it must be connected with the choice of technique in relation to the discrete economic context of a particular sector. Having said this, I should also emphasize that the choice of any particular technique does not in itself reveal enough about "the meaning or end to which the action is ultimately oriented" to permit an accurate assessment of the presence or absence of modern rational capitalism. Such an assessment requires additional data about the cultural context and the different "ends" of social action currently prevailing, as well as some understanding of the general relationship of the economy to society.

A second qualification concerning the rationality and modernity of the Italian Renaissance economy relates to its sectoral concentration on commerce and banking as compared with large-scale manufacturing. This emphasis, which involved a far greater stress on circulating than on fixed-capital investment, placed considerable limits on the extent to which bookkeeping methods involved a fundamental rationalizing and methodizing of business life. The economic rationality of Italian Renaissance did not extend, for example, to a rigid separation of "income" from "capital" or to the practice of periodic valuation and revaluation of productive possessions as capital assets (Yamey 1950, pp. 23–24; Pollard 1972, pp. 122–24).

This has an important implication for those theorists like Sombart, who place accounting methods in such a central position in their theories of capitalist development. For while Sombart may be correct in arguing that the technique of double entry is a precondition for the development of the idea of capital as an impersonal productive entity separate from the businessman (Sombart 1919, 2:120), there is no evidence to suggest that this conception of capital, so crucial to the modern epoch of heavy fixed-capital investments, had emerged with the development of double entry in the Renaissance.

This in turn suggests the need for caution in interpreting Renaissance accounting simply in terms of a gradual evolution toward the present. There is, after all, no historic certainty that the accounting innovations mentioned by Cohen led inevitably to the modern idea of capital as an impersonal yet productive resource. Instead they may have coexisted over a long period with otherwise traditional business institutions based ultimately on considerations of kinship and status rather than economic rationality.

Cohen's exaggeration of the scale of rational capitalism (in Weber's sense) within Renaissance Italy depends finally on an inadequate consideration of the relationship between economy and the wider polity and social order. In this respect insufficient attention is given to the extent to which economic activity, however much it embodied commerce and profit seeking, was nonetheless still largely embedded within a social setting that did not set the rational and continuous pursuit of profit above all else. Cohen rightly draws attention to some indicators of this situation, notably, the lack of free markets (Cohen 1980, p. 1347) in commodities and to a lesser extent labor, the infant state of rational technology (Cohen 1980, p. 1349), and governmental restriction on economic activity (Cohen 1980, p. 1350). However, he fails to weigh the restrictive and limiting implications of such phenomena for economic activity, in particular, for the extent to which they offset the need for businessmen to act under the imperatives of continuous rational profit seeking.

In general terms, while Renaissance entrepreneurs strove to enlarge their incomes and fortunes, there is considerable evidence, even amid the sources Cohen utilizes, that they neither wished nor needed to pursue profit in a rational and continuous manner as an overriding principle of conduct. Given that markets were still regulated, the legal-political situation was unstable and incalculable, and communications were slow, business success did not depend extensively on rational cost accounting (even in the few cases where this was practiced) or on free market efficiency. It was based, rather, on servicing the political, fiscal, and luxury consumption needs of a relatively small elite dominated by royal and noble households, the papacy, and the urban patriciate (de Roover 1956, p. 124; Hale 1979, p. 13). Such needs were satisfied in the main by banking facilities and international trade in high-quality goods. With the exception of textiles, manufacturing played a far less central role in Italian Renaissance economic success. Even here the emphasis from the 14th century onward was on the luxury end of the market, not on any move toward standardized production for a mass market. Such activity, while indicative of economic expansion, fails to meet an important criterion set by Weber for the presence of modern capitalism, namely, "the provision of everyday wants by capitalistic methods" (Weber 1961, p. 207).

While some individual merchants (Lane 1967; Origo 1957, p. 95) did take enthusiastically to the role of entrepreneur, this was rarely their exclusive or predominant concern. Especially once economic success was secured, traditional status considerations such as the use of wealth to marry into the nobility or an older family or to achieve influence within the city-state or papacy (Lane 1966, p. 38; Hale 1979, pp. 13–15) became of more enduring significance. The rise to dominance of the Medici family within the civic politics of Florence, culminating in its transformation

from a republic to a hereditary dukedom, indicates a far more intense and sustained preoccupation with status elevation than with rational banking and commerce, on which the family's fortunes were originally built.

This testifies in part to a shift from the role of entrepreneur to what Weber called the role of "rentier and occasional entrepreneur" (Weber 1978, 12:1295) among the larger mercantile families of this period. Sometimes the shift occurred in less than one generation. The same process is reflected in the general preference for investing mercantile fortunes not in manufacturing industry but in landed estates. Such status aspirations have caused Hale, the historian of Renaissance Florence, to speak of a re-feudalization process during the 15th century, which Cohen regards as a high point in the development of rational capitalism. This was characterized by an emphasis on "country villas, chivalrous poems, tournaments, an interest in genealogy," betraying furthermore "the attractiveness to wealthy merchants and bankers of an exclusiveness based on blood rather than effort" (Hale 1979, p. 81).

Underlying such phenomena was the continuing close relationship among business, kinship, and the large family household. Although both Weber and Cohen have drawn attention to the beginnings of a separation between kinship ties and economic activity (Weber 1961, pp. 172–73; Cohen 1980, p. 1344), there were still considerable limits to its development. Certainly, the separation of business and personal accounts had begun, while the practice of taking nonfamily partners in a venture was widespread (Weber 1961, pp. 171–72; Origo 1957, pp. 101–2; Lane 1953, p. 91; de Roover 1963, pp. 43–52). Nonetheless, on the basis of existing research it is still unclear whether the importance of the family and patriarchal authority was declining within Italian Renaissance business. De Roover (1948, p. 34; 1963, p. 47), for example, shows how nonfamily partners in the Medici bank became incorporated into the Medici household by marriage.

Finally, Cohen seems to be under a misconception when he describes Renaissance business houses as "to some extent . . . like Weberian bureaucracies" (Cohen 1980, p. 1343). This is not just a matter of the continuing importance of the ascriptive ties of kinship compared with individual recruitment and achievement based on merit. It also concerns the preference for incorporating additional manpower in banking and mercantile operations through partnership in capital rather than salaried status (Lane 1966, p. 43–44; de Roover 1963, pp. 43–52, 80; Origo 1957, pp. 101–2). There is also the question of scale. While Cohen (1980, p. 1343) pictures the Medici enterprise as involving "bureaucratic chains of command includ[ing] senior partners, a general manager, branch managers, assistant managers, purchasing agents and bookkeepers, shop fore-

men, and laborers," de Roover, historian of the Medici bank, presents a quite contrary analysis. "It would be a grievous mistake," he argues (de Roover 1963, p. 43), to have an exaggerated idea of the size of the medieval banks and to picture them as giant institutions doing business in office buildings with marble lobbies and rows of windows behind which a crowd of employees operate machines and manipulate papers. In reality, "medieval bankers transacted their business in small offices adorned with a bank or counter, a few desks, and in the rear the abacus on which the bookkeeper makes his computations by casting his counters."

CONCLUSION ·

Cohen's attempt to rebut Weber's emphasis on the Reformation as a decisive contributor to the spirit of modern capitalism, by reference to rational capitalism in Renaissance Italy, seems both conceptually misconceived and empirically exaggerated. Engagement with Weber demands a more elaborated discussion of the concepts of rationality and modern capitalism than Cohen supposes. But it also requires a far more sensitive handling and evaluation of historical evidence as presented in the monographic literature. Fruitful intellectual dialogue between sociology (conceptualized by Weber as the formulation of "type concepts" and search for general uniformities) and history (viewed as the causal analysis of individual actions, structures, and personalities that have cultural significance) (Weber 1978, 1:19) is evidently far more difficult to secure than Cohen's article suggests.

REFERENCES

Cipolla, Carlo. 1967. *Clocks and Culture, 1300–1700*. London: Collins.
Cohen, Jere. 1980. "Rational Capitalism in Renaissance Italy." *American Journal of Sociology* 85 (6): 1340–55.
de Roover, Raymond. 1948. *Money, Banking and Credit in Medieval Bruges*. Cambridge, Mass.: Medieval Academy of America.
———. 1956. "The Development of Accounting prior to Luca Pacioli according to the Account Books of Medieval Merchants." Pp. 144–74 in *Studies in the History of Accounting*, edited by A. C. Littleton and Basil Yamey. Homewood, Ill.: Irwin.
———. 1963. *The Rise and Decline of the Medici Bank*. Cambridge, Mass.: Harvard University Press.
Ehrenberg, Richard. (1928) 1963. *Capital and Finance in the Age of the Renaissance*. Reprints of Economic Classics. New York: Kelley.
Hale, J. R. 1979. *Florence and the Medici*. London: Thames & London.
Lane, Frederick C. 1953. "Family Partnerships and Joint Ventures in the Venetian Republic." Pp. 86–101 in *Enterprise and Secular Change*, edited by Frederick C. Lane and Jelle Riemersma. Homewood, Ill.: Irwin.
———. 1966. *Venice and History*. Baltimore: Johns Hopkins University Press.
———. 1967. *Andrea Barbarigo, Merchant of Venice: 1418–49*. New York: Octagon.
Le Goff, J. 1960. "Temps de l'Eglise et temps du marchand." *Annales économies, sociétés, civilisations* 15:417–33.

Luzzatto, Gino. 1953. "Small and Great Merchants in the Italian Cities of the Renaissance."
Pp. 41–52 in *Enterprise and Secular Change,* edited by Frederick C. Lane and Jelle
Riemersma. Homewood, Ill.: Irwin.

Origo, Iris. 1957. *The Merchant of Prato.* Oxford: Alden.

Parsons, Talcott. 1949. *The Structure of Social Action.* 2 vols. New York: Free Press.

Pollard, Sidney. 1972. "Capital Accounting in the Industrial Revolution." Pp. 119–44 in
Capital Formation in the Industrial Revolution, edited by François Crouzet. London:
Methuen.

Roth, Guenther. 1976. "History and Sociology in the Work of Max Weber." *British Journal
of Sociology* 27 (3): 306–18.

Sapori, Armando. 1953. "The Culture of the Medieval Italian Merchant." Pp. 53–65 in
Enterprise and Secular Change, edited by Frederick C. Lane and Jelle Riemersma.
Homewood, Ill.: Irwin.

————. 1969. "The Physiognomy of the Florentine Merchant." Pp. 65–76 in *Social and
Economic Foundations of the Italian Renaissance,* edited by Anthony Molho. New York:
Wiley.

Sombart, Werner. (1915) 1967. *The Quintessence of Capitalism.* New York: Fertig.

————. 1919. *Der moderne Kapitalismus.* 2 vols. Munich: Humblot & Duncker.

Tawney, R. H. 1926. *Religion and the Rise of Capitalism.* New York: Harcourt, Brace.

Thompson, E. P. 1974. "Time, Work-Discipline and Industrial Capitalism." In *Essays in
Social History,* edited by M. W. Flinn and T. C. Smout. Oxford: Clarendon.

Weber, Max. 1924. "Zur Geschichte der Handelsgesellschaften im Mittelalter." Pp. 312–443
in *Gesammelte Aufsätze zur Sozial- und Wirtschaftsgeschichte.* Tübingen: Mohr.

————. 1958. *The Protestant Ethic and the Spirit of Capitalism.* Translated by Talcott
Parsons. New York: Scribner.

————. 1961. *General Economic History.* Glencoe, Ill.: Free Press.

————. 1978. *Economy and Society.* 2 vols. Berkeley and Los Angeles: University of Cal-
ifornia Press.

Yamey, Basil. 1950. "Scientific Bookkeeping and the Rise of Capitalism." In *Studies in
Accounting,* edited by W. T. Baxter. Sydney: Law Books of Australia.

Reply to Holton[1]

Jere Cohen
University of Maryland Baltimore County

R. J. Holton, in his critique of my (1980) article "Rational Capitalism in Renaissance Italy," has challenged my evidence on five specific points: whether Italian Renaissance capitalists used time efficiently or wasted it; whether standardized products were being developed to satisfy an abstract circle of purchasers; the extent to which double-entry bookkeeping was employed; the extent to which the trial balance was used regularly; and whether the firms were like Weberian bureaucracies.

On the first point, I do rely, as Holton says, on the opinion of Armando Sapori (1969), who, as Holton concedes, "clearly has a detailed knowledge of business records." Similarly, de Roover (1941, p. 29) notes that "the duties of the entrepreneur laid a claim on all his time." The efficient use of time had not yet been raised to a guiding principle of conduct as it was in Benjamin Franklin's writings; this came much later. However, the medieval Catholic church preached in favor of the wise use of time.

Holton argues that my first two references concerning the development of standardized goods for an abstractly conceived market give my position little support. However, since Ehrenberg ([1928] 1963, p. 55) argued that true commodity exchanges require that "the commodities must be so far standardized that they do not have to be seen at the time of the transaction," the presence of Renaissance commodity exchanges implies some standardization of goods. When Holton read Ehrenberg's opinion that there were no true commodity exchanges in the Middle Ages, he apparently assumed that this applied to the Renaissance as well, which would have invalidated my argument. However, Ehrenberg ([1928] 1963, pp. 233–37) reports the development of commodity exchanges in Bruges and Antwerp by 1484 and states that "many commodities were now made in standard types." Regarding my second reference, Holton seems to have missed the support given by de Roover (1963, p. 144) in reporting the manufacture of cheap tapestries with repeating patterns that "could be cut to fit the shape of any room, like wallpaper today." This production for the market contrasted with the manufacture of the more expensive tapestries to fill special orders. Holton argues that since the final buyers

---

of cloth were often unknown, the goods were shipped without a buyer in mind, and it was uncertain whether the goods would sell, there was no "abstract circle of purchasers within the businessman's calculations." However, de Roover (1948, pp. 41–42) makes it clear elsewhere that sellers shipping goods abroad generally acted on advance information about scarcity and price. The pre-estimation of risks in shipping has also been described by Beard (1962) and Einstein (1913). The Levantine cloth market, for example, was fairly well known in Renaissance times: at least one Florentine company put a special finish on cloths to cater to the Levantine taste (see de Roover 1941, p. 19). This definitely shows that Italian merchants had some concept of their market. It is true that de Roover considered business an adventure insofar as goods were shipped first and a buyer searched for later. This occurred because, at that time, the manufacturers of goods were also responsible for their sale to consumers and needed to ship goods abroad to be sold by their foreign agents. However, to the extent that Renaissance merchants became full time and sedentary, remained in business for decades, repeated operations, and utilized regularly scheduled shipping (see Cohen 1980, pp. 1342–43), I must agree in part with Postan (1952) that commerce had "ceased to become an adventure." What seems most likely is the coexistence of some rational capitalism and some adventure capitalism, which Weber (1958, p. 21) believed occurred "even in modern Western countries."

Holton questions whether double-entry bookkeeping was in general use in Renaissance Italy. Yet, according to Lane (1967, p. 158), "it seems reasonable to conclude that double entry was well-known and commonly used at Venice at the beginning of the fifteenth century" and "probably well-known and commonly used throughout Italy." Although Lane (1967) knew of only four Venetian account books, these *all* used double entry, and Florence was ahead of Venice (Peragallo 1938). The Medici firms used double-entry bookkeeping, and according to de Roover (1941, p. 25), "the general accounting procedure found in the Medici account books . . . correspond[s] to the general practice of the time as described in all the existing works on the history of accounting." Luca Pacioli, who had described the double-entry method in 1494, said that he was writing down the system that had been used in Venice for over 200 years (see Taylor 1956, p. 180); in consistence with this, both Peragallo (1938, p. 73) and de Roover (1941, p. 28) concur that "bookkeeping practices then current in business were far superior to treatises on the subject." Holton has calculated from de Roover' s (1956) data that half of all post-1390 account books were kept in double entry, and even if his 50% figure were accurate, I would still judge that this constituted "general use"; however, the 50% figure is in all probability low because (1) he has counted incomplete and therefore indeterminate record fragments as negative cases and (2) most

of the account books cited in that article are not even from Italy! De Roover (1956) surveyed a number of other European countries, too, all behind Italy in accounting procedures. Moreover, Holton is incorrect in assuming that de Roover's (1956) article includes "virtually all extant account books of the Italian Renaissance," since a number of different books are analyzed elsewhere (e.g., Peragallo 1938). Holton is probably correct that the smaller businesses lagged behind the larger in accounting practice, as is true today. In support of this point, he presents Origo's (1957) quotation from Datini stating that the small-town Prato merchants reckoned their accounts in their heads rather than in books; on the same page, Origo contrasts the situation in the cities: "Most of the great Florentine companies, however, as well as the Venetians and Genoese, already used the system of bookkeeping by double entry. . . ." Weber (1961, pp. 170–74), though he knew that the Italians had contributed to bookkeeping, did not credit them with the development of *double-entry* bookkeeping (see Weber 1961, p. 212).

I have claimed general use for the trial balance, but not that it was used *regularly* (i.e., at regular intervals) by the majority of firms. Balancing the books to close a ledger is not considered a true trial balance, but the trial balance was widely known and usually computed whenever a partnership was dissolved or a partner died or withdrew; there are numerous instances of this, not just three or four. Thus, Peragallo (1938, p. 31) seems correct when he states that "the use of the balance sheet was general." A different and less general usage was the repeated but *irregular* use of trial balances. One example of this is in the firm of Andrea Barbarigo, which struck a trial balance every five or 10 years (see Lane 1967). Another example is the striking of a balance to determine assets and liabilities for tax purposes; many Florentine firms submitted balance sheets for the *catasti* (municipal tax assessments) of 1427, 1433, 1451, and 1458 (see Lane 1967, p. 92). Holton is correct that there are only a few definite instances of an *annual* trial balance; I can name only the Medici companies and the Borromei, Datini, and del Bene firms. Since, however, there were a *number* of Medici companies in Florence, most of which balanced annually, a significant percentage of the Florentine cases known to historians used an annual trial balance. De Roover (1956, p. 157) was willing to generalize that "the Florentine companies drew up balance sheets every year, or, at least, at more or less regular intervals. . . ." In contrast, trial balances were used infrequently in Venice (Lane 1967, p. 181); as Holton says, they made little sense for merchants in the long-term international trading ventures that were typical of Venice. A contemporary observer, Pacioli (cited by Peragallo 1956, p. 217), may have summed up matters best when he wrote in 1494 that "leading merchants" in "famous places" balanced the ledger every year. This is an important point. "What is at

stake here" is more than Holton realized, more than debits, credits, or managerial control in a wide sense: without double-entry bookkeeping, a proper balance cannot be struck; without a proper balance, the businessman cannot calculate profit or loss; and without this knowledge, business decisions cannot be based on the calculation of their profitability. Because this type of decision-making rationality is central to Weber's (1961, p. 207) concept of modern capitalism, it is important to show the use of double-entry bookkeeping and of the trial balance. While the former seems firmly established in 1500, the regular use of the latter was not as widespread.

Fifth, and last, Holton denies that Italian firms were bureaucracies, saying they were too small; but there is no contradiction between relatively small size and a bureaucratic structure. The same de Roover who attests to the small size of the banks presents organizational charts for the Medici bank (1948, p. 12) and the Rafaello di Francesco de' Medici Company (1941, p. 21) that show a complicated formal division of labor and a hierarchy of authority. The bureaucratic chains of command that Holton considers contrary to de Roover are, *in fact, described by de Roover* on the pages cited (see Cohen 1980, p. 1343); de Roover (1941, pp. 11–18) has described an even more complex operation than I did. Besides small size, Holton sees ascription as compromising the bureaucratic nature of these firms. It is true that relatives were taken into the Medici firms, but they were dropped if they did not work out well (e.g., de Roover 1948, p. 25; de Roover 1956, pp. 43–52). According to de Roover (1956, p. 52), the importance of family ties was on the wane in business (although they were still important socially).

As the foregoing demonstrates amply, my "handling of the empirical data" *is not* "in certain key instances, seriously inaccurate," and hence my assessment of rational capitalism in Renaissance Italy is not "empirically exaggerated." The best available historical opinion backs up all my (1980) claims.

Holton reminds us quite usefully that historians' judgments are often based on small and unrepresentative numbers of cases: for example, in the present debate we have no random sample of Renaissance firms and no time-budget studies of Italian entrepreneurs. If we are to use the highest standards of sociological evidence, we cannot be secure in accepting most of the generalizations made by historians. For this reason, the dialogue between sociology and history is, as Holton concludes, indeed a difficult one. However, the implications of this problem go beyond Holton's conclusions and could ultimately damage rather than support Weber's thesis. In order to be scientific, Weber's thesis must be capable of disconfirmation by some obtainable evidence; conversely, it needs historical evidence to *support* it. A historical thesis like Weber's needs to be tested through

historical evidence if it is to be tested at all. Otherwise, it is not a scientific hypothesis but a fairy tale about a faraway ancient land that we can never know. By the standard of evidence necessarily employed to test Weber's thesis, the Italian Renaissance data are quite good.

Beyond questioning the evidence itself, Holton questions whether it shows the presence of rational capitalism in Weber's sense. According to Weber (1961, p. 207), "a rational capitalistic establishment is one with capital accounting, that is, an establishment which determines its income yielding power by calculation according to the methods of modern book-keeping and the striking of a balance." Weber believed that the balance was developed around 1700, but by his own criteria there were rational capitalistic establishments in Renaissance Italy. Weber (1958, p. 21) defined capitalism as the ever-renewed pursuit of profit through continuous enterprise, and I (1980) have demonstrated the profit orientation, repeated operations, sedentary nature, and relative stability that typified Renaissance Italian firms. In addition, Weber (1958, pp. 21–24) presented a list of the "peculiarities of Western capitalism" that made it very different from the forms of capitalism that appeared elsewhere; besides "rational book-keeping," this list consisted of "the rational capitalistic organization of (formally) free labour," rational industrial organization attuned to a regular market, and "the separation of business from the household." When I showed that these characteristics as well as most of Weber's "presuppositions" of modern capitalism were present in some degree in Renaissance Italy, I showed the presence of "rational capitalism" in a sense that (1) is taken directly from Weber, (2) is systematic rather than "loose," and (3) is as complete as Weber's lists of rational capitalism's characteristics and presuppositions.

Holton states that my demonstration of rational capitalism in Weber's sense is incomplete because it lacks evidence that businessmen pursued profit "as an end in itself" and as "a guiding principle of life"; but he confuses "rational capitalism" with the *spirit* of capitalism, which Weber (1958, p. 51) described as "above all the idea of a duty of the individual toward the increase of his capital, which is assumed as an end in itself." I have found no evidence that the "spirit of capitalism" existed in Renaissance Italy. However, Weber (1958, pp. 64–68) maintained that the capitalistic form of enterprise and the spirit of capitalism may occur separately from each other. Hence, it is possible within Weber's conceptual framework to show the presence of rational capitalistic enterprise even in the absence of the "spirit of capitalism"; that was precisely the case in Renaissance Italy. For example, the Medici bank engaged in continuous rational profit seeking for decades at the same time that Lorenzo de' Medici maintained broader interests in politics and the arts.

Despite its lack of an *ethic* of profit seeking, Renaissance capitalism

did not lack a "spirit"; its spirit was not traditional in nature but, as I (1980) showed, was a spirit of rational calculation. Weber apparently considered this "rational spirit" separate from "a rationalistic economic ethic" since he (1961, p. 260) listed them separately; yet he (1958, p. 55) felt that a "peculiarly calculating sort of profit seeking" was closely related to the spirit of capitalism. The Italian example shows that capitalism can develop a spirit of rational calculation without an ethical duty to increase capital.

In Holton's view, modern capitalism in Weber's sense does not exist unless capitalism predominates to such an extent that without it the economic system would collapse; I did not apply this criterion to Renaissance capitalism because I (1980) was not claiming that the "whole epoch" was capitalistic (see Weber 1961, p. 207). Capitalist development was still incomplete during the Renaissance: while there was *some* "provision of everyday wants by capitalistic methods" (e.g., textiles) (Weber 1961, p. 207), capitalism still coexisted with more traditional economic forms and did not yet predominate.

Although Roman Catholics did not produce a complete and dominant system of rational capitalism, they did contribute to the early development of rational capitalism, and their contributions were important links in the chain of events that eventually (though not inevitably) led to its completion and dominance two centuries later. The problem with Weber's ideal-type method is its sharp dichotomy between (1) an epoch where rational capitalism predominates and (2) one where it exists but could be removed without an economic collapse. This distinction minimizes the early contributions of Catholics to rational capitalism since it emphasizes their lack of a dominant capitalist system; conversely, by dating the shift to capitalism as having happened after the Reformation, it maximizes the apparent Protestant contribution. However, the idea of a discrete post-Reformation shift masks the long-term gradual growth of rational capitalism, which occurred both before and after the Reformation. The early contributions to rational capitalism, which helped to form it, were as significant in their way as the later contributions that made it dominant; and both Protestants and Catholics were able to advance capitalist rationality significantly.

Despite Weber's obvious "historical sense" and his (1961) awareness of pre-Reformation "antecedents" and "forerunners" of rational capitalism, he clearly underestimated the development of such rational elements as the capitalist firm, double-entry bookkeeping, the trial balance, and the rational spirit of calculation in pre-Reformation times. These aspects of modern capitalism were able to develop considerably without the inclusion of the Protestant ethic or the spirit of capitalism among their causal antecedents.

REFERENCES

Beard, Miriam. 1962. *A History of Business*. Ann Arbor, Mich.: Ann Arbor Paperbacks.
Cohen, Jere. 1980. "Rational Capitalism in Renaissance Italy." *American Journal of Sociology* 85:1340–55.
de Roover, Raymond. 1941. "A Florentine Firm of Cloth Manufacturers." *Speculum* 16:3–33.
————. 1948. *The Medici Bank*. New York: New York University Press.
————. 1956. "The Development of Accounting prior to Luca Pacioli according to the Account Books of Medieval Merchants." Pp. 144–74 in *Studies in the History of Accounting*, edited by A. C. Littleton and Basil Yamey. Homewood, Ill.: Irwin.
————. 1963. *The Rise and Fall of the Medici Bank*. Cambridge, Mass.: Harvard University Press.
Ehrenberg, Richard. (1928) 1963. *Capital and Finance in the Age of the Renaissance*. Reprints of Economic Classics. New York: Kelley.
Einstein, Lewis. 1913. *The Italian Renaissance in England*. New York: Columbia University Press.
Holton, J. R. 1983. "Max Weber, 'Rational Capitalism,' and Renaissance Italy: A Critique of Cohen." In this issue.
Lane, Frederick C. 1967. *Andrea Barbarigo, Merchant of Venice: 1418–1449*. New York: Octagon.
Origo, Iris. 1957. *The Merchant of Prato*. Oxford: Alden.
Peragallo, Edward. 1938. *Origin and Evolution of Double Entry Bookkeeping*. New York: American Institute Publishing Co.
————. 1956. "Origin of the Trial Balance." Pp. 215–22 in *Studies in the History of Accounting*, edited by A. C. Littleton and Basil Yamey. Homewood, Ill.: Irwin.
Postan, Michael. 1952. "The Trade of Medieval Europe: The North." Pp. 119–256 in *The Cambridge Economic History of Europe*, vol. 2. Edited by Michael Postan and E. E. Rich. Cambridge: Cambridge University Press.
Sapori, Armando. 1969. "The Physiognomy of the Florentine Merchant." Pp. 65–76 in *Social and Economic Foundations of the Italian Renaissance*, edited by Anthony Molho. New York: Wiley.
Taylor, R. Emmett. 1956. "Luca Pacioli." Pp. 175–84 in *Studies in the History of Accounting*, edited by A. C. Littleton and Basil Yamey. Homewood, Ill.: Irwin.
Weber, Max. 1958. *The Protestant Ethic and the Spirit of Capitalism*. Translated by Talcott Parsons. New York: Scribner.
————. 1961. *General Economic History*. Glencoe, Ill.: Free Press.

Commentary and Debate

To conserve space for the publication of original contributions to scholarship, the comments in this section must be limited to brief critiques. They are expected to address specific errors or flaws in articles and reviews published in the *AJS*. Comments on articles are not to exceed 1,500 words, those on reviews 750 words. Longer or less narrowly focused critiques should be submitted as articles. Authors of articles and reviews are invited to reply to comments, keeping their replies to the length of the specific comment. The *AJS* does not publish commenters' rebuttals to authors' replies. We reserve the right to reject inappropriate or excessively minor comments.

ETHNIC TIES AND POLITICAL MOBILIZATION:
COMMENT ON LEIFER

Eric Leifer ("Competing Models of Political Mobilization: The Role of Ethnic Ties," *AJS* 87 [July 1981]: 23–47) has contributed several potentially useful additions to the growing empirical literature on the persistence of ethnic ties in modern states. Leifer points out accurately that the developmental and cultural division of labor models share the prediction that ethnic attachments will continue to affect political outcomes under cultural division of labor conditions. He also specifies correctly that an accurate test of the cultural division of labor model includes an interaction term between economic underdevelopment and cultural subordination. He also argues that the role of leaders within ethnic movements should be taken directly into account when the impact of ethnic attachments in political mobilization is analyzed. However, none of these observations is new (see Ragin 1979, p. 631; Nielsen 1980, p. 78; Mughan and McAllister 1981, p. 201). Several shortcomings in the theoretical exposition and data analysis, however, may be responsible for the absence of significant findings in the data. Since Leifer interprets these results as evidence that "ethnic ties serve as a facilitating condition for political mobilization" (p. 44), it is important to clarify these issues.

Though Leifer claims to present a "crucial" test between competing

theories (p. 24), his regrettable omission of ethnic competition theory leads him to present data which merely recount what is already known. For example, using similar data, Ragin presented evidence in 1977 that Labour party vote is unrelated to peripheral sectionalism when considered alongside support for Conservative party vote. Leifer also dismisses ethnic competition predictions in note 5, even though Ragin (1979) and Mughan and McAllister (1981), using voting data from Great Britain, document considerable support for ethnic competition hypotheses (see also Nielsen's [1980] research on the Flemish Nationalist party vote in Belgium and Olzak [1982] on Parti Québécois vote in Quebec).

Given the strong evidence presented in both Ragin's and Nielsen's work, it is not clear why Leifer chose to measure the impact of ethnic ties on political mobilization using Conservative, Liberal, and Labour party vote as dependent variables. Ragin (1979), for example, presents data showing clear differences in sources of Liberal and Welsh Nationalist party support, concluding that analysis of the Welsh Nationalist party is a more direct indicator of ethnic attachments than is analysis of the dominant British political parties. Leifer has chosen to ignore the significant historical and content differences between parties, which leads to some confusion over the magnitude and direction of effects we should expect in his data. I argue that the confusion arises early in Leifer's paper because of the focus on "political mobilization" rather than "ethnic political mobilization"; the latter is the central concern of the developmental, cultural division of labor, and ethnic competition models. The difference in focus between the two is that emphasis on political mobilization in general obscures the persistent role of ethnic organization (when compared with class, occupational, or some other basis), whereas analysis of ethnic political mobilization considers ethnic party strength directly in competition with other possible attachments.

Implicit in Leifer's definition of ethnic attachments is the largely discredited developmental theory definition of ethnicity as "primordial" (see Enloe 1973; Hechter 1975; Young 1976; Yancey, Ericksen, and Juliani 1976). His argument regarding the role of ethnic ties in modern political systems ultimately rests on the persistence of ethnic characteristics which "define a common discourse . . . which is required for any form of concerted action" (pp. 28–29). This definition is problematic: consider those cases in which ethnic politics are divorced from specific cultural content (e.g., as Furniss [1979] argues in the case of Scotland) or cultural revivals in language or tradition result from political organization (see Linz 1973; McRoberts 1979). Part of this problem can be traced to Leifer's omission of resource mobilization explanations of political organization (see McCarthy and Zald 1977; Tilly 1978). For example, ethnic competition theories link indices of modernization and development with ethnic mo-

bilization by arguing that, when development increases the stock of re-
sources (economic, political, demographic, or some other resource) available
to a formerly disadvantaged group, ethnic competition can result in ethnic
group mobilization in the political sphere (Nagel and Olzak 1982). This
latter formulation allows consideration of regions undergoing ethnic
movements in states where core-periphery boundaries are less clear-cut
(see Mughan and McAllister [1981] for an example of this in Great Britain).

The omission of the politically constructed nature of political mobili-
zation in Leifer's paper is particularly puzzling. Modern states create
ethnicity in a variety of ways: through the expansion of the political sector
in forming bureaucratic offices and channels for ethnic grievances, through
designation of certain ethnic groups as targets of civil rights programs,
and by generally expanding the scope of political competition to include
formerly unincorporated groups (the category of "American Indians," forged
from many tribes in the United States, is an important case in point,
according to Nagel [1979]). Also, it is not clear from Leifer's discussion
of political mobilization in Great Britain what role the state has played
in creating or co-opting ethnic groups over time. In addition, there is the
possibility that a supranational organization (such as the EEC) has both
a direct and an indirect impact on subnational ethnic organization, through
various policies concerning immigration of guest workers, international
price controls, and regional inequality, to name but a few (Nagel and
Olzak 1982).

A final point concerns Leifer's attempted explanation for the absence
of significant effects found for both the cultural division of labor and
developmental models. Leifer argues that developmental and cultural
division of labor models omit a "means-specific leadership component"
which could be measured by educational attainment of cultural minorities
in both core and peripheral educational institutions (p. 45). He argues
further that the location of such leaders' education may determine whether
an ethnic movement becomes integrated into the dominant national party
system, or whether ethnic movements become oriented toward separatist
goals: "For example, a leadership component that directs mobilization
toward the national political arena is generally developed within the
educational institutions of the dominant ethnic group. Members of the
subordinate ethnic group trained in these institutions have a tendency to
direct their political energies back toward the national 'center' in seeking
political change. A counter logic can be applied to those trained in local
centers, who are more likely to press for 'separatist' political platforms"
(p. 45).

Although there may be disagreement about what constitutes core or
local educational institutions, there is well-documented evidence refuting
Leifer that shows that the majority of independence, subnational, and

separatist movement leaders receive their education in institutions dominated by the core ethnic group or in colonial centers abroad (Bell 1964; Wallerstein 1965; Young 1976; McKowan 1974).

In addition, as Young (1976) and others have pointed out, an increase in the proportion or number of subnational leaders from the periphery is *not* independent of the economic and political resources mobilized in an ethnic periphery. It is only when an ethnic periphery becomes partially integrated into the national political network that educated leaders mobilize movements. Leifer needs to consider that his "leadership component" variable is inextricably linked both to conditions of ethnic subordination and to subsequent economic development in ethnic regions. More to the point, leadership (and the rise in resource mobilization potential which it indicates) can arise only when economic and political conditions *change* the relationship between core and periphery. Models like Leifer's, which analyze the static "facilitating conditions" of ethnic political mobilization, therefore miss the important dynamic quality implied by political mobilization processes.

By omitting consideration of the dynamic processes inherent in political mobilization, Leifer has seriously weakened his contribution to the empirical literature on the role of ethnic attachments in modern party structures. By focusing on the relationship between characteristics of peripheral regions and dominant party vote in Great Britain, he loses sight of what substantive difference ethnic attachments may make in modern political party contests. Inclusion of ethnic competition models would have led him to consider changes in ethnic mobilization potential, which would include measures of the leadership potential in the periphery, as both causes and consequences of political mobilization. In consideration of the current state of the art and the problems in Leifer's work, these tasks remain ahead for researchers of ethnic mobilization in modern states.

SUSAN OLZAK

Yale University

REFERENCES

Bell, Wendell. 1964. *Jamaican Leaders*. Berkeley and Los Angeles: University of California Press.
Enloe, Cynthia. 1973. *Ethnic Conflict and Political Development*. Boston: Little, Brown.
Furniss, Norman. 1979. "The Political Component of Scottish Nationality." Pp. 152–79 in *Ethnic Autonomy—Comparative Dynamics*, edited by Raymond T. Hall. New York: Pergamon.
Hechter, Michael. 1975. *Internal Colonialism: The Celtic Fringe in British National Development, 1536–1966*. London: Routledge & Kegan Paul.
Linz, Juan. 1973. "Early State-building and Late Peripheral Nationalism against the State: The Case of Spain." Pp. 32–116 in *Building States and Nations*. Vol. 2, edited by S. N. Eisenstadt and Stein Rokkan. Beverly Hills, Calif.: Sage.

McCarthy, John, and Mayer Zald. 1977. "Resource Mobilization in Social Movements: A Partial Theory." *American Journal of Sociology* 82 (May): 1212–39.

McKowan, Roberta. 1974. "The Impact of Education on Ethnic Groups in Ghana and Kenya." Pp. 233–54 in *Ethnicity and Nation Building,* edited by Wendell Bell and Walter Freeman. Beverly Hills, Calif.: Sage.

McRoberts, Kenneth. 1979. "Internal Colonialism: The Case of Quebec." *Ethnic and Racial Studies* 2 (July): 293–318.

Mughan, Anthony, and Ian McAllister. 1981. "The Mobilization of the Ethnic Vote: A Thesis with Some Scottish and Welsh Evidence." *Ethnic and Racial Studies* 4 (April): 189–202.

Nagel, Joane. 1979. "The Political Construction of Ethnicity." Paper presented at the annual meeting of the American Sociological Association, Boston, August.

Nagel, Joane, and Susan Olzak. 1982. "Ethnic Mobilization in New and Old States: An Extension of the Competition Model." *Social Problems* 30 (December): 127–43.

Nielsen, François. 1980. "The Flemish Movement in Belgium after World War II: A Dynamic Analysis." *American Sociological Review* 45 (February): 76–94.

Olzak, Susan. 1982. "Ethnic Mobilization in Quebec." *Ethnic and Racial Studies* 5 (July): 253–75.

Ragin, Charles. 1977. "Class, Status, and 'Reactive Ethnic Cleavages': The Social Bases of Political Regionalism." *American Sociological Review* 42 (June): 438–50.

———. 1979. "Ethnic Political Mobilization: The Welsh Case." *American Sociological Review* 44 (August): 619–35.

Tilly, Charles. 1978. *From Mobilization to Revolution.* Boston: Addison-Wesley.

Wallerstein, Immanuel. 1965. "Ethnicity and National Integration in West Africa." Pp. 474–88 in *Africa, Social Problems of Change and Conflict,* edited by Pierre van den Berghe. San Francisco: Chandler.

Yancey, William, Eugene Ericksen, and Richard Juliani. 1976. "Emergent Ethnicity: A Review and Reformulation." *American Sociological Review* 41 (June): 391–403.

Young, Crawford. 1976. *The Politics of Cultural Pluralism.* Madison: University of Wisconsin Press.

CONTEXT AND CONTENT IN MODELS OF POLITICAL MOBILIZATION: REPLY TO OLZAK

Failing to find anything new or promising in my article, Susan Olzak has offered us an alternative framework. Though I too was frustrated by the negative results of my research, I could justify remaining within the reactive ethnicity framework by my discovery of a logically necessary component that was missing in my own, and all prior, formulations: a means-specific leadership component. My framework focuses exclusively on context, while Olzak's tries to incorporate content, in examining the origins of ethnic-based political mobilization. I will compare these alternative frameworks to help stimulate "crucial tests" between them (though the different focuses make such tests highly problematic). First, however, I will review some of my backward-looking contributions, since Olzak tends to treat them as "confusions" rather than as bases of a framework very different from her own.

Rather than treat the *content* of interests, in terms of class or status

(ethnic) orientations, I treated ethnic commonalities as a facilitating condition (*context*) for mobilization. The shift from content (as a dependent variable) to context (as an independent variable) had the following implications:

1. The focus of the developmental (DEV) and reactive ethnicity (RE) models shifted from the rhetoric associated with political activity to the degree of mobilization that will occur in concrete settings. Distinguishing between class and ethnic interests in a cultural division of labor is a futile exercise because of the necessary overlap between these interests. When ethnic ties are treated as facilitating conditions for mobilization, however (and mobilization is taken as problematic), the two models differ in their quantitative predictions, since the DEV model affords no role for ethnic ties. The fate of both models comes to rest on the significance of a single interaction coefficient, which is the *only* basis for distinguishing between them.

2. The contextual framework allowed the unambiguous use of aggregate data: contexts can have clear geographic boundaries; interests cannot be pinned down so easily.

3. The shift away from the content of interests removes the historical component from the framework, as is essential for generality. Analysis of party platforms is no longer needed; only the degree of support for any direction pursued is needed (see below).

The focus I use is consistent with Hechter's (1975) verbal formulation, but Hechter went astray in introducing Weber's distinction between class and status interests to prop up his analysis of residuals method (the use of residuals would require an *empirical* independence between class and status interests—which cannot hold in a cultural division of labor. The analytical independence in Weber's distinction means nothing to the regression grinder). Ragin (1977) recognized the lack of independence between class and status interests in a cultural division of labor but went right ahead and used Hechter's form of residual analysis. Ragin (1979) also introduced an interaction term but failed to recognize it as the basis for a crucial test. Instead, he formulated separate equations for each model and sprinkled interaction terms in each of them. This left him in the position of comparing a lot of coefficients outside an unambiguous hypothesis test format. Olzak would not have had such an easy time rejecting my formulation on empirical grounds had I merely presented the "same data" in the same form as those before me. This is my backward-looking achievement. I will now try to deal with the more forward-looking orientation of Olzak's comments.

Olzak seems to be advocating an amalgamation of the ethnic competition (EC) and resource mobilization (RM) models, combined with a

sensitivity for the role of the state and the "politically constructed nature of political mobilization." My less sprawling suggestion (see pp. 44–46) was to incorporate the contextual implications of a missing leadership component into the predictions of the RE model. It is useful to understand the relations between these alternative directions.

First, consider the way Olzak and I partition our subject and the populations we study. Olzak insists that I should study "ethnic political mobilization," while I make it clear from title to conclusion that I am interested in "political mobilization." Ethnic ties are a residual phenomenon in the class-oriented DEV model and, thus, could hardly be "the central concern," as she claims. The role of ethnic ties is acknowledged by the RE model under the special condition of a cultural division of labor, but otherwise the DEV and RE models do not differ. *Only* in the EC model of Olzak are ethnic ties the invariant central concern. The EC model lacks the apparatus to treat class (or regional economic—see my n. 6) conflict; it treats only ethnic conflict *within* status positions. This is why I had trouble formally comparing the EC model with the other two (see my n. 5), especially in the context of aggregate data. It is also why I have trouble understanding Olzak's accusation that I treat ethnicity as "primordial," in light of my accusation that the EC model treats ethnicity as "fundamental" (n. 5). Ethnicity, in my framework, refers to any commonality (outside of interests) that facilitates interpersonal communication. It becomes "reactivated" when it is perceived as the basis for discrimination within the stratification system.

Differences in the partitioning of the subject are reflected in the choice of a dependent variable. In my framework of political mobilization, it is necessary to use a dependent variable that is measurable in all regions of the political system. Separatist party vote poses problems here, since it is by nature limited to specific regions. The problem is easily solved, however, by using the *concentration* of vote for *any* party within each region instead of the percentage of vote for a particular party as the dependent variable. This would be the most precise measure of political mobilization (within the framework of democratic elections). It was not really necessary in my study because separatist parties failed to win a significant portion of votes in the period I studied. This approach to the dependent variable stands in sharp contrast to Olzak's approach, with its interest in "indicator[s] of ethnic attachments" and "significant historical and content differences between parties." The shift from focusing on the content of interests to the context of mobilization is responsible for this sharp contrast. In my framework, an optimal dependent variable is one that can indicate concerted mobilization, *whatever its direction may be.*

This lack of interest in the direction of mobilization may seem in conflict with the incorporation of a missing means-specific leadership component. One is tempted to argue that if a completely general indicator for degree of mobilization (of any kind) were found, it would no longer be necessary to worry about specific directions of mobilized effort. But this ignores the necessary role of direction-giving leadership in mobilization, in that such leadership is a *resource* of widely varying nature that can be present or absent from concrete contexts. It must be present in some form (along with the other necessary conditions) to lead us to expect any form of concerted mobilization; it must be present in a particular form to lead us to expect a particular kind of mobilization (I was interested in the latter case, since I was using national election data). Clearly, we need to know more about this leadership component (I thank Olzak for correcting my substantive conjecture—it was intended only as an illustration of how to incorporate the component, in whatever form). The fact that the leadership component may be interdependent with the nature of the core-periphery relationship, as Olzak points out, does not affect this logic. Interdependence affects only the conditional distributions of contextual properties (i.e., the size of the regions in the Venn diagram on p. 45) and not the effects we attribute to a given context. My treatment of leadership as a resource should have satisfied Olzak—it is an application of the resource mobilization logic that does not leave the specification of "resources" to a post hoc inventory but instead pins down the motivational and directional resources necessary for concerted mobilization.

It is on the topics of leadership and the possible incorporation of the role of the state that the boundary between the two alternative directions. is the most problematic. In my formulation, all the components were contextual properties; that fact facilitated the unambiguous use of aggregate data. Yet treating the missing leadership component as reducible to a contextual property should be regarded as a bit forced. Context never fully determines orientations—value indeterminacies are an integral part of the action framework. Yet if we reject this contextual treatment, and take the strategies of leadership (and the state) as given data within a gaming framework, we must be prepared to face the myriad methodological ramifications that these new elements of content introduce. The shift from "static" (contextual) models to "dynamic" models involves more than the introduction of more "-tion's" into our rhetoric. It is one thing to advocate "dynamic" models and another thing to come up with one. The sophisticated discontent of Olzak may be a bit premature in this research area.

ERIC M. LEIFER

Harvard University

Review Symposium

Theoretical Logic in Sociology. Vol. 1: *Positivism, Presuppositions, and Current Controversies.* By Jeffrey C. Alexander. Berkeley and Los Angeles: University of California Press, 1982. Pp. xx + 234. $25.00.

THE ALEXANDER QUARTET: EARLY IMPRESSIONS[1]

Gianfranco Poggi
University of Edinburgh

Positivism, Presuppositions, and Current Controversies is the first volume of what is to be a staggeringly large work. Three more volumes, each (I am told) longer than this one, are to follow. The publication of the four volumes will complete a project Jeffrey Alexander began while still a Berkeley graduate student under the supervision of Robert N. Bellah and Neil Smelser. The book is, in fact, an augmented and rewritten version of Alexander's huge, much praised 1976 doctoral dissertation.

From the beginning, the project was as remarkable in its ambitiousness as its product is in size because Alexander consciously set out to produce nothing less than *The Structure of Social Action (SSA)* of his generation. Like *SSA, Theoretical Logic in Sociology* seeks to (re)establish the very foundations of sociology by locating and solving the discipline's most general and consequential theoretical problems. A further, conspicuous analogy with Parsons's 1937 book lies in the strategy central to both works, the pursuit of its aim by means of a close, critical reconstruction of the major theoretical legacies from the discipline's past. These are, in Alexander's judgment, the legacies of Marx and Durkheim (to be discussed in vol. 2), of Max Weber (vol. 3), and of Parsons himself (vol. 4). There is no indication, however, that Alexander intends to ground his argument on anything similar to Parsons's dubious claim that the theories of his own principals—Alfred Marshall, Durkheim, Pareto, and Weber—*converged* toward a single conception of "the structure of social action."

Thematically, Alexander's first volume corresponds roughly to the first three chapters of *SSA* in that it aims to establish the legitimacy of the undertaking, to articulate its methodology, and to outline some of its results. In substantive terms, the analogy with *SSA* is particularly strong in Alexander's first chapter, "Theoretical Logic in Scientific Thought." Here he vigorously justifies a sustained concern with those components of scientific work that are not connected wtih observation and empirical

generalization; using Parsons's own terms, he argues that theory consti-
tutes "not only a dependent but an independent variable in the devel-
opment of science" (*SSA*, p. 6). In this context, Alexander makes excellent
use of the considerable advances, mostly since World War II, toward a
"postpositivist" understanding of the scientific process, and in particular
the awareness that, even in the natural sciences, theory is underdeter-
mined by facts.

Early in chapter 1, Alexander diagrams a "continuum of scientific
thought" (p. 3) along which he ranges a number of distinctive components
according to their degree of generality and their proximity to what he
calls the "metaphysical environment" of the scientific process. In the next
chapter he uses this continuum to point up deficiencies in previous at-
tempts to map out the most general tendencies and controversies within
sociological theory. Thus, "mappings" such as those of Habermas, which
align theories on one side or the other of the controversy about positivism,
are given a low generality rating by Alexander because of their prevalent
concern with methodological matters. Those (e.g., Dahrendorf's) that
focus on "equilibrium vs. conflict" (an odd formulation, this one) do not
get a much higher rating because that issue, as Alexander sees it, reduces
to the empirical question of whether there is more conflict or more equi-
librium in a given society. Even the attempts by Mannheim or Gouldner
to characterize social theories primarily by their politico-ideological im-
plications fall short—according to Alexander—of "those fundamental
conceptual issues in sociological thought that do constitute, by virtue of
their generality and decisiveness, the truly presuppositional level of so-
ciological theory" (p. 63).

While the discussion summarized above has no close parallel in *SSA*,
the thematical analogy becomes compelling once again in chapter 3. Here
Alexander argues that two terms central to Parsons's argument—"action"
and "order"—do indeed belong at that "truly presuppositional level,"
makes extensive use of other Parsonian expressions ("individualism," "ra-
tionality," "normative," etc.), and employs sources (chiefly Halévy) much
used by Parsons. He also follows Parsons's lead on a number of substantive
points. Sometimes he, as it were, out-Parsons him: for instance, in his
discussion of the problem of action, Alexander nails himself firmly to the
"two-dimensional position" which Parsons himself subscribed to but which
he ultimately surrendered—a position according to which action has nec-
essarily both an external reference to constraining, mostly material con-
ditions, and an internal reference to ideal, subjective commitments.

If I make so much of the Parsonian motifs in this chapter—the central,
as well as the longest, in the volume—it is not in order to play down the
considerable originality of Alexander's effort. Instead, it is because, in
this chapter, when Alexander more or less follows Parsons, I seem to
know what he is doing; when he does not, I do not. I am not clear, for
example, how he can argue that all major sociological theories support
the "two-dimensional position" on the problem of action, and then give
a variety of contrasting solutions to the same problem. Alexander's dis-

tinction between the "epistemological" and the properly sociological level of analysis, which is meant to help us overcome that difficulty, does not do so because it is hurriedly and inadequately explicated. Nor do I understand why, if the problem of action and of order are both as "truly presuppositional" and at the same time as conceptually different as Alexander (in opposition to Parsons) insists, only the former problem makes its appearance both at the epistemological and at the sociological level.

Perhaps the difficulties I have with this chapter, in spite of its containing a number of clear and significant points, come from Alexander's (or his publisher's) decision not to publish all volumes of the book at the same time. In *SSA*, arguments comparable with those advanced by Alexander in chapter 3 are presented as the validated results of Parsons's contextual discussion of Marshall, Pareto, Durkheim, and Weber; thus the reader can satisfy himself whether or not that discussion validates those arguments. Alexander's chapter 3 is severed from the discussion of *his* principles, and he does not make it clear whether its validity is dependent on that discussion. If it does, then send in the clowns, and we shall see. If it does not, then it is as yet unclear why Alexander, after his presentation of otherwise validated conclusions, expects his readers to move on to the later volumes. Confronted only with volume 1, I will not judge this four-course meal until all of it has been served.

I do have a few preliminary comments on what readers may find in this volume and, presumably, expect in the forthcoming ones. I believe they will find that Alexander is a thoroughly competent scholar: his sources range very widely and his treatment of them, insofar as I can judge, is both penetrating and fair; some of his numerous footnotes are even impressive micro-essays in their own right. I am less impressed by the quality of his writing. The arrangement of his arguments is clear and persuasive; but the exposition, while generally effective, is often ponderous and never brilliant. Not enough of his sentences are as striking as this one: "Metaphysics can be no more eliminated from science than the non-rational element can be eliminated from action" (p. 17); and too many are as clumsy as the following: "The social scientific problems that are truly generalized, that are neither positivistic nor conflationary, have been identified" (p. 112).

A large number of infelicities and minor errors may be unavoidable in a work of this size. For instance: In the well-known metaphor, the "force" is not "unmovable" (as he puts it) but "irresistible" (p. 189, n. 34). "Correlary" (p. 207, n. 144), "en toto" (p. 203, n. 135), "nonvaledictory" (p. 193, n. 70), "indeterminant" (p. 62) are lexicological innovations we can do without. On pages 181 (n. 4), 204 (n. 136), and 209 (n. 151), respectively, "consolation," "inculcation," and "injunction" are misused. There is no "Social Studies Unit" at Edinburgh University (p. 133). It is mischievous to suggest that Percy Cohen follows Alexander rather than vice versa (p. 68). I object to Alexander's extravagant overuse of the term "logic"; it crops up all over the place and in God knows how many unspecified meanings.

Finally, let me suggest a simple test of how quickly and widely this book will get read and talked about. During the next few years, just listen for the expressions "conflationary" and "presuppositional" in the talk of your colleagues and students. If Alexander's massive and ambitious work has the expected impact, the frequency of these will rise sharply.

Parsons, Jr.[1]

Alan Sica
University of Kansas

Epigraph Florilegium:

> *Ancient:* . . . it collects cinnamon, and spikenard, and myrrh, and of these materials builds a pile on which it deposits itself, and dying, breathes out its last breath amidst odor. From the body of the parent bird, a young Phoenix issues forth, destined to live as long a life as its predecessor [500 years]. [OVID A.D. 7]

> *Radical:* Historical moments visit twice, first as tragedy, then as farce. [KARL MARX, approximately (1852)]

> *Popular:* And here I sit so patiently,
>   Waiting to find out what price
> You have to pay to get out of
>   Going through all these things TWICE.
>   [BOB DYLAN 1966]

Writings which open with a fusillade of the past's pithy remnants sometimes irritate more than entice. Not only may they appear pretentious or hesitant in disclosing their own position openly, they may also seem to indulge in a clubby assumption that all readers will join the author in taking from the precious quotations an identical message. I do not want to risk baffling anyone yet, so let me be plain about Jeffrey Alexander's mighty project. Social theory has been here before, first in 1937 when Parsons published *The Structure of Social Action* (which won a tiny readership [Parsons (1937) 1968, p. v]), then in 1959 when Mills dissected "Grand Theory" in his enduring manifesto (pp. 25–49). Later on Sorokin had a go at ruining Parsons (1966, pp. 403–40), but with less success than Mills because of his intemperate charge that Parsons had plagiarized him (highlighted in the famous "similarity tables" of alleged thefts), and be-

[1] I wish to thank Robert J. Antonio for his critical comments.

cause Sorokin himself was an aged grand theorist. Of course, legions have attacked Parsons's lifework from every available angle, and he responded sometimes directly but more often obliquely to the continuous pounding. Yet like all those who attempt innovation, for better or worse, he believed he was right and never gave up much terrain. The stubborn continuity of his ideas and their expression seems astounding, almost heroic, when considered against the attacks they sustained for 40 years.

This became particularly clear to a younger generation of sociologists when, during the mid-1970s, extensive reappropriation of the classics began, with "de-Parsonizing" one of its central projects. His defense against these determined assaults was no less obstinate, especially regarding his interpretations of Weber and Durkheim, than were his earlier rejoinders to critics (notably the debate over his conception of power and coercion within "equilibrium" states). Still in 1976 he felt "enhanced confidence" in the theory of convergence that had been dismantled by dozens of scholars over the decades (Parsons 1976, p. 364). Yet soon after his death in 1979, as Bierstedt observes (1981, pp. xii, 394), the sharp question Parsons had borrowed from Crane Brinton to open his career, "Who now reads Spencer?" haunted his own ghost: "Who now reads Parsons?" Soon no one will need agonize over *The Social System* again, for instead, they can turn to Alexander's four volumes *if* structural-functionalism suits them.

Giddens and others believe that Parsons's first book was his best, mainly because it brought new European thought to American sociology and the former has since dominated the theories of the latter. Alexander does not shower exotic material on his audience. Instead he has tried to purify Parsons's approach, then protect it from all comers. He is a ruthless lieutenant in his general's tent, ferreting out the disloyal and dismissing them on charges of "conflationary reasoning," which for Alexander means unredeemable stupidity. Alexander never conflates; he inflates. Monumental, enormous, and tremendous theoretical breakthroughs, fashioned either by Parsons, his older followers (most often Smelser, Barber, Eisenstadt, Bellah, and Lipset), or Alexander himself, occur at every turn. Absolutism reigns; you are either on the bus, like Kesey's merry pranksters, or you are theoretically retrograde. And Alexander is driving. He has wrested the wheel from Smelser and the other elders—or perhaps they slid to the back of the bus to make room for him—and now they shout his praises dutifully, observing him drive the straight, narrow Benthamite path set by Parsons in 1937. For Lipset the book is "brilliant and original"; Barber plays Brutus: "We hail a new master sociologist"; and Smelser, who codirected the dissertation that became the book, becomes prophet: "As future thinkers view 20th century sociology they will without

doubt regard Jeffrey Alexander's [book] as the most important effort in theoretical synthesis of its generation." Perhaps future thinkers—should there be a future for thinkers—will treat sociology as an involuted fad that ossified into self-parody around 1982, then languished. Or merged with action painting. Or perhaps they will agree with Smelser, while noting that only Alexander was naive enough, so late in the game, to try another "theoretical synthesis"—the holy grail of the field—so he "wins" by default.

This is not to say that his efforts over the past decade are worthless or that his reading of some social theory is not provocative. Rather, studying Alexander's first volume, along with his other work (1976, 1977, 1978, 1980, 1981a, 1981b, 1982 [hereafter, unless indicated otherwise, all citations are to works by Alexander]), puts one grimly on the slope beside Sisyphus. Mills wrote, "Just now, among social scientists, there is widespread uneasiness, both intellectual and moral, about the direction their chosen studies seem to be taking . . . it is this condition, of uneasiness and indifference, that is the signal feature of our period" (1959, pp. 19, 12). And he was clear about the job at hand: "It is now the social scientist's foremost political and intellectual task—for here the two coincide—to make clear the elements of contemporary uneasiness and indifference [in] . . . this general malaise of contemporary intellectual life" (pp. 13, 19). He could be writing today. Mills and others pushed the stone Parsons had carved to the crest of the hill and into the light where it teetered for two decades, the elements wearing it away. Yet, once again, almost in disbelief, we find ourselves in the valley, this time beneath a clone of the functionalist megalith. It is more coyly shaped than its progenitor's, louder in its autocritique, but under the new getup the same assumptions about theorizing prevail. Confidence in relative societal homogeneity; theoretical remoteness from subjectivity, even from genuine interest in the living subject, coupled with rhetorical intimacy about freedom and choice; the familiar, deadening Parsonian tic, to squeeze whatever comes to hand within omnivorous, fetishized Concepts (as Mills observed); all are there. This means, of course, that the library of critiques—from Bierstedt's early blasts (1938a, 1938b) to the latest de-Parsonizing—must be taken from the shelves again, or, more suitably, enlarged a bit to fit the new wrinkles. This is the discouraging, Sisyphean prospect.

Alexander's basic thrust was displayed in his first major article (1978). Before subscribing at $100 to the four-volume *Theoretical Logic* set, it would be rational to "study" (as Barber insists we must) Alexander's celebration of Parsons in that cheaper vehicle. There one sees all the felicities and vulgarities of his method, of what he views as proper theorizing. The tone of *Positivism, Presuppositions, and Current Controversies* is somewhat more genteel, as one would expect from a volume supplied, like the Bible, with

a ribbon bookmark. But Alexander's strategy is continuous and reveals itself quite plainly. A selective list of his rhetorical tropes, his hermeneutical ploys, might introduce us efficiently to his project.

The Problem of Hubris

Except for Parsons ("a great thinker"; 1978, p. 177) and Halévy ([1901–4] 1972), the historian of utilitarianism, no one escapes strident censure. To be sure, a theorist needs self-confidence in spinning out the new line. But reliable guides for finding that happy balance between fawning adoration and caustic dismissal of predecessors have not been provided for students of social theory. We have no accepted hermeneutic method, as do scholars in philosophy, literature, law, and theology. In addition to this basic deficiency, social theorizing, until very recently, has been the preserve of egomaniacs, a provincial exclusionary Men's Club. (We know that Weber subscribed to Durkheim's journal but refused to mention him, and that Parsons breathed hardly a word about Marx in the mid-1930s when everyone else was in thrall to him.) But when a theorist also poses as an intellectual historian, he or she must partake of a humbling historicist corrective in pursuing sound readings, which, after all, is what Alexander is doing most of the time. Good readership, practiced hermeneutics, is nowadays much more crucial, I think, to responsible theorizing than before, particularly when Parsons was young. It simply cannot be carried out when hindered by haughty reductionism and trivialization. One must now see the "rupture" in the seam of received wisdom that is being proposed as itself thoroughly conditioned—by private ignorance, the zeitgeist, muted political dreams, and so on. What can one make, then, of a young man (born 1947) who begins by announcing *ex cathedra* that Lockwood, Coser, Dahrendorf, Mills, Martindale, Gouldner, Rex, Foss, Friedrichs, Bottomore, Levy, Smelser, Mayhew, Wallace, Rocher, and six others were all mistaken about Parsons because "none of these descriptions [as opposed to analyses?] . . . is sufficiently generalized to comprehend Parsons's most fundamental theoretical contribution" (1978, p. 178)? Later he writes, in a note, that "Habermas has failed throughout most of his career to make significant progress on any of these questions" (1978, pp. 193–94n.). A favorite phrase, "Contrary to conventional wisdom . . ." (e.g., 1978, p. 180), crops up repeatedly.

In disappointing contrast to the bluntness of Sorokin or the graceful clarity of MacIver, for example, Alexander's voice seems congenitally oracular. His 1978 article begins, in fact, with a Ciceronian eulogy to Parsons: "Only with the passage of time, as the center loses its immediate power, can a perspective which is both critical and appreciative be attained and the thinker's permanent contributions to intellectual tradition

be properly assessed" (p. 177). There is never a coming down to earth. Even in analyzing Watergate in a later paper, Alexander cannot risk appearing comfortably mundane: "I began this essay with a rather esoteric theoretical problem and concluded it with a detailed case study of a gaudy and famous political scandal" (1982, p. 24). He wants badly to be accepted as a serious theorist, full of esoteric ways, who penetrates to the core of social arrangements and brings light to dark places. And he reminds his readers of this every so often, lest they mistake him for a regular person. In another paper he maligns Habermas for having a Germanic, elitist "fear of the consequences of independent empirical insight" (1981b, p. 287); at the same time he himself sides with the laity's "moral rationality" (p. 286). He wants to have it both ways—to stand at the pinnacle of social thought, pronouncing truth about differentiation and the dynamics of substantive voluntarism, at the same time remaining a true American, a man of the people. He has taken over many of Parsons's mannerisms, and this is one of the most costly. He is doing social theory in the upper stratosphere, while claiming his feet are firmly on the ground, like just plain folks'.

There is nothing wrong with headstrong pride in a social theorist; the same can be said of the concert virtuoso. But when virtually every scholar or thinker beyond the tiny elect circle of functionalists—and the former group includes scores in *Theoretical Logic*—is belittled, the reader's sympathies gradually turn from the attacker to the attacked. Put another way, if Alexander's goal were to articulate as economically and convincingly as possible a new or revised theory of social action, or just fresh methodology prior to such a theory, the proper strategy would be to bare it simply, free of the incessant disparaging of others that mars both the *ASR* article and the book under review.[2] Scholarship and theorizing are either communal enterprises, based on judicious—*not* slavish—appro-

---

[2] So many unmistakable, negative similarities obtain between Parsons and Alexander that I can hardly do better than quote Bierstedt on Parsons, since most of this passage applies to the younger man as well: "Two other mannerisms may briefly be mentioned. The first . . . concerns what Devereux calls Parsons's 'exasperating tendency to insist that each and every point in his entire system is fundamental.' Another favorite word is *crucial*. . . . The second mannerism is a puzzling one. In person and in personal relationships Parsons was the most modest of men. Unassuming, uncritical of others, reluctant to mention himself in conversation, much less to talk of his accomplishments, he becomes the opposite in writing. In writing he claims for himself many 'major breakthroughs' . . . he equates his own theory with sociological theory. . . . He has a tendency, in short, to usurp the function of the critic. . . . He writes indeed as if he were his own principal disciple" (1981, pp. 394–95). Joseph Gabel, the psychiatrist and social theorist, has written at length on such anti-dialectical or "de-dialecticized" relations to the social world. He generalizes from his clinical findings to presumably normal ideological and political adversaries, who propagate monological discourse in the belief that their views are uniquely correct, thus beyond external critique (Gabel 1978).

priation of past achievements, past sallies into newness, or else we are monads, reinventing time, space, and Being during each grant period.

The Problem of Language

Bierstedt (1981, pp. 393–95) begins his entrancing chapter on Parsons with "The Great Language Barrier," the finest, most succinct critique of Parsonese since Mills's. After quoting an impenetrable sentence, he notes, "Although there are, of course, lapses into clarity, this is all too typical an example of his prose" (p. 393). Alexander writes English less badly than his inspirer, but he does carry on the functionalist tradition of semantic legerdemain, particularly when he finds himself at a tendentious point. There is first, as noted above, the apodictic, Mosaic tone throughout, for example, "The originality of this aspect of Parsons's contribution, which has been thoroughly misconstrued by interpreters within the individualist tradition, is, then, his comprehension of the essential compatibility of an emphasis on voluntary individual will with a collectivist emphasis on normative interpenetration" (1978, p. 180). This says that Parsons discovered people do what they want as far as social rules allow.[3] As Alexander sees it, Parsons is always "original"; he has been mercilessly "misconstrued" until now; and Parsons, unlike anyone else, finds "essential compatibility" between opposites, like voluntarism and societal constraint.[4] Like so many of Alexander's domain assumptions, all this hints at kinship with Milton Friedman's *Free to Choose*.

The "individualist tradition" according to Alexander began within "strands of Enlightenment thought": "These are the traditions which

---

[3] Parsons did not discover this, though he thought its primitive formulation lay in Hobbes. When Alexander mentions Hobbes (1982a, pp. 98 f.), he takes his information from Parsons, quoting the same passage from *Leviathan*. Parsons found his fabled "problem of order" on pp. 24, 43, and 63–66 in the Everyman edition of Hobbes's masterpiece. Alexander says elsewhere, "Parsons was actually quite ruthless with his predecessors . . ." (1981, p. 505), which is one of his few understatements. But this anticipates a problem on the use of sources, to be dealt with below.

[4] Alexander sometimes likens Parsons to Hegel (e.g., 1978, p. 179) in the former's joyful willingness to bring together conceptually what life as normally experienced keeps separate. The hint is that Hegel's pivotal *Widerspruch* ("contradiction") somehow resembles Parsons's persistent fusing of, say, idealist, materialist, and individualist "traditions" into his own "unified" action theory (1978, pp. 185, 189). (In one essay Alexander prophesies that changes within Hegelian thought owing to distinct "national traditions" may be echoed in different national schools of Parsonianism [1981a, p. 501].) Without going into detail, I want to note that Parsons practices crude Kantianism, strewing artificially resolved antinomies throughout his work. Hegel's treatment of *Aufhebung*, especially in the smaller *Logic*, is worlds removed from Parsons's clumsy, undialectical procedures, as even casual acquaintance with the texts reveals. Alexander also points to a logical similarity between Parsons's "elements of action" and Husserl's levels of phenomenological reduction (1978, p. 182), which I find absurd, and parallel to the alignment of Parsons and Hegel. This is sheer rhetoric.

stand at the heart of nineteenth century liberal ideology and emphasize free will as the principal ethical criterion of freedom. In its social scientific form, this ideological point is transposed into a distinctively individualistic theoretical position, which perceives social action as initiated by, and society as resting upon, discrete individuals who are free to pursue their interest as they have defined it" (1978, p. 179). This means that Bentham believed people should do what they can get away with; most of the words in this passage are superfluous, wrong by simplification of the record, or both. Alexander cites Martindale's (1960) text, his section on Kantianism, and Halévy's ([1901–4] 1972) old standard as sources for his reductions. He then says that symbolic interactionism, exchange theory, and phenomenological and existentialist sociologies all share this heritage, since they apotheosize (my choice of word) the individual. This, too, is wrong, first because these schools did not evolve from a monolithic "individualist tradition," whatever that is, but from various philosophical and social-psychological streams which do not fit neatly on an individualist-collectivist continuum; and second, because they do not uniformly begin with "the freedom of the individual person" as their a priori basis of theorizing. So, having caricatured the Enlightenment and its "individualist strands"—without any inkling that Adorno and Horkheimer's *Dialectic of Enlightenment* shattered the comfortable Carl Becker version to which he still clings—and having misgrouped current schools of theory, Alexander administers the *coup de grace* to his straw figures:

> Parsons's response to this tradition [Alexander is forever pointing out traditions unknown until he names them] argues that its voluntarism is based upon a radical misunderstanding of the theoretical role of the concept "individual," a problem which he attributes to a metamethodological problem; namely, the empiricist confusion of concrete and analytic frames of reference. . . . It is at this point that Parsons arrives at his great insight into the voluntary quality of action. He reasons that if no individual can actually be free of constraint in the radical sense propounded by individualist theory, then what we normally perceive as free, intentional activity must in fact involve the actor's application of an internal normative standard of judgment. [1978, p. 180]

I am quoting at length because these sentiments lie at the root of Alexander's theoretical and axiological weltanschauung; if you do not accept them, his edifice falls. This paragraph argues that Parsons believed he alone had recognized solipsism for what it is, an adolescent fancy, a robinsonade of the type steamily portrayed in Sartre's *Nausea*. And this is a "great insight"? Should Sumner have called his book *Internal Normative Standards of Judgment* instead of *Folkways?* Must every shred of sociological knowledge be attributed to Parsons's sanitizing and ordering obsession?

I did not choose these paragraphs for a brief hermeneutic because they are egregious; far from it. They are quintessential Alexander. The game is one of words, not referents (of semantics and syntactics, not imaginative correspondence to reality, as Mills noted long ago [1959, pp. 33–35]). First he asserts that everyone writing before Parsons needed and received correction from his master. Then he points to the stunning, unique revelations penned by Parsons and follows up with capsule denunciations of those who have since misinterpreted him. After this, in culmination, Alexander presents his own "refinement" of Parsons's ideas. Among the problems with this relentless procedure are these: Parsons was not a sophisticated student of philosophy or epistemology, as shown, for example, by his dubious handling of idealism and emanationism in 1937. If he had read Brentano, Husserl, Scheler, Heidegger, Wittgenstein, or Peirce, he concealed it. Much of his understanding of 19th-century thought derived from Halévy, thus the heavy Benthamite influence. His readings of Pareto, Weber, and Durkheim, and his remarkable exclusion of Marx, have all been called into question. Yet Alexander holds onto the "theoretic-epistemic" raiments of Parsons's first book as if no critiques had been published, or those that had were worthless. He is in love with a vocabulary and the veils it brings its user.

The celebrated historian of global dimension, William H. McNeill, reviewed in these pages a book by a member of Parsons's inner circle, S. N. Eisenstadt, a comparativist who figures very importantly in Alexander's work as a sort of functionalist Barrington Moore. This is what McNeill said about Eisenstadt's *Revolution and the Transformation of Societies:*

> It seems to me that the argument is largely logomachy. S. N. Eisenstadt first defines true revolution in such a fashion that his analysis only fits events in early modern Europe and then treats the implication of his definition, elaborately worked out in dreary polysyllabic prose, as a great discovery. . . . I find little to admire in such word games and deplore three qualities of the book. First and foremost, it is written in a private jargon that is hard for the uninitiated to penetrate. . . . Second, Eisenstadt uses his private language in a way that is so general as to be well-nigh meaningless. . . . Third, I feel that Eisenstadt is a dogmatist at heart. . . . [His] polysyllables have not exhausted the truth about revolutions, nor, I think, done anything to illumine my understanding of them. [1980, pp. 1241–42]

A thorough analysis of Alexander's rhetoric, formally speaking, would take many pages, but the end result would coincide, I believe, with McNeill's (1980) opinion of Eisenstadt's creative functionalist semantics. We still forgive Derrida (and his followers) for pages which make little sense, because the suspicion lingers that his revolution in language will abet a positive revolution in critical thought. Clearly Alexander continues

to believe that functionalist language does good things, that it transcends competing jargons and gets to the root of scientific knowing. It is equally clear that not many people beyond his own small encampment agree with him. It becomes a Foucaultian battleground, of power dictating knowledge. Alexander chastised Habermas for despising empirical inquiry while idealizing dialogical communication. But if a theorist listens only to the cheering sounds of his faithful, and touches the empirical only when pressed, he does indeed play at logorrheic logomachy, the last word in reified "thought."

Before continuing with objections to Alexander's theorizing, I want to summarize my charges thus far, for they could be perceived as "presubstantive," therefore inconsequential. Style is indeed the person, and though I have waited to deal directly with some of his ideas, shorn of rhetoric and subterfuge, I consider the problems already reviewed equally important in evaluating his innovations. *Theoretical Logic* seems to me slavishly Parsonian, though slyly so, since the master's wrist is slapped now and again, but never severely. The scholarship is suspect, given to rampant "conflation" of schools, periods, methods, ideas. Real care is taken only in assessing Parsons's own words, and some of his students'. Others are dismissed or damned with faint praise. He bullies the opposition, the unfaithful, which results, as it always has, in weak hermeneutics. Weber knew how to read his enemies fairly (see esp. *Roscher and Knies* [1975]) while advancing his own theories; it can be done. Alexander, like Parsons, has a tin ear for language, and also like Parsons, seems to have no affinity for the play of culture.[5]

Science Hypostatized

A return to the mid-1960s when the National Science Foundation thought sociologists were scientists, and behaved accordingly, would please many people. Alexander tries reviving the period early in his book by sounding out Toulmin, Koyré, Polanyi, Lakatos, Kuhn, Gerald Holton, and others in his section "Theoretical Logic in Scientific Thought." He believes there is such a thing and argues vigorously against Kuhn, who "erred by conflating the different levels of generality on the scientific continuum" (p. 25), thereby risking a "narrowing and reductive impact on the analysis of science" (p. 27). Alexander dislikes anything "anti-intellectual and ir-

[5] This may seem like a whimsical complaint, but recall that the first books of Mannheim and Lukács, for example, seemed so stimulating owing partly to the authors' training in art history, drama history, and aesthetics. Parsons came out of biology, then economics, and was never able to illuminate the arts from within his theory. This results in mechanical thinking and stiff prose. In all of this one might see irrelevance, but *only* if the model for social theory were taken from Newton rather than Hegel, say, or William James.

rationalist"—considers "extreme subjectivity" (p. 83) the death knell of theorizing—and speaks up for battered "commitments to 'rationalist' notions of evidence and logic; more generally, commitments to shared disciplinary conceptions of proper scientific goals" (p. 25). It frightens him, as well it would any grand theorist, that some things, important ones, may not be orderable, such as the ways a scientist's brain works, or the reasons he or she has for doing whatever is done and afterward called "science." A friend of Kuhn cannot be a friend of Alexander, so the latter also berates Robert Friedrichs's history of recent sociology, which "seriously misrepresented the nature of the different elements of sociological thought . . . because of his 'conflationary' approach" (p. 26).[6] Alexander feels particularly compelled to set Friedrichs straight because the latter said so many unkind things about Parsons. But he does not admit this; instead, rather typically, he argues a series of needlessly thorny points, never mentioning the most likely cause of his consternation (pp. 26, 43, 149).

Alexander gets down to business by offering four principles on which, he claims, all "postpositivist" philosophers of science agree. (Shades of Parsons's convergence theory surface, noticeably so when he converges Lakatos and Feyerabend. Yet this packaging rankles less here because he is trying to advance a position more than to write intellectual history.) They are (1) all scientific data are theoretically informed, (2) empirical commitments are not based solely on experimental evidence, (3) general theoretical elaboration is normally dogmatic and horizontal rather than skeptical and vertical, and (4) fundamental shifts in scientific belief occur only when empirical changes are matched by the availability of alternative theoretical commitments (pp. 30–33). In this part of the book one must put the ribbon bookmark to vigorous use, as we are instructed on page 1, since endnotes 118–44 pertain to only three text pages, exhausting six pages and two dozen authorities. Perhaps I missed something monumental, but I do not know why this kind of defensive scholarship is necessary to give these four principles credence, unless Alexander does not really believe in his own convergence: "That among those theorists I have described as postpositivist there often exists disagreement should, indeed, be emphasized. Some writers exhibit strains toward idealism, others toward empiricism. Nonetheless, in terms of the specific criteria I have set out, I believe their work can be classified under this common framework" (p. 154, n. 118). The Parsonian dodge is again at work. In 1937 the story was that four men with nothing manifestly in common were actually working out, unknown to themselves, "the immanent development of the

---

[6] Perhaps in the future graduate students in sociology—should there be any—can be inoculated with a serum against "conflationary paralysis."

logic of theoretical systems in relation to empirical fact," with the "voluntaristic theory of action" as its fruit (Parsons [1937] 1968, p. 14). Alexander's analogue:

> Despite the wide substantive disagreement as well as analytic confusion among the various advocates of the postpositivist position, there are certain basic points upon which all agree and which present a perspective on science that is radically at odds with the one shared by the sociological protagonists of the positivist persuasion [William Catton, Hans Zetterberg, William Goode, Homans, sometimes Merton]. . . . The fundamental basis of this agreement can be most effectively articulated in terms of the issue which in one form or another has preoccupied every thinker I have mentioned, namely, the relation between "theory" and "fact." According to the representatives of the postpositivist persuasion, *all scientific development is a two-tiered process, propelled as much by theoretical as by empirical argument.* [P. 30, emphasis in original]

I do not understand why this is momentous. To begin with, how many sociologists today emulate Zetterberg, Catton, or Homans? Why does Alexander believe his counterposition is shocking or relevant to social research? Perhaps because he perpetuates Parsons's refusals to look around him, such as, "For the purposes of this study—not necessarily for others—it is legitimate to define philosophy as a residual category" (Parsons [1937] 1968, p. 43). Is it ever "legitimate" to ignore philosophy when theorizing about epistemology, even humble versions suited to social science? It is hard to believe that Alexander could be as ignorant of as much contemporary thinking about science, epistemology, and the logic of inquiry as he pretends; that he can seriously believe today that "idealism" and "empiricism" describe an exhaustive continuum of positions—especially after his own lengthy review of Richard Bernstein's esteemed text *The Restructuring of Social and Political Theory* (1978).

Alexander insists that "science proceeds as surely by a generalizing or 'theoretical logic' as it does by the empirical logic of experiment" (p. 33). Facts, theories, pseudofacts, and crypto-theories are involved in dialectical interplay, as their definitions, their specifications change with generations, political arrangements, and economic cycles. (Foucault, absent from Alexander's citations, has commented on this.) But from these agreeable remarks Alexander reaches the ledge and tries to jump to another level: "If the nature of social science is to be properly understood, and its true potential fully achieved, the careful attention to methodological rules for induction from empirical observation must be matched by an effort to create a 'theoretical methodology' that can explain the opposing movement from more general principles" (p. 33).

This reads as if cut from the same cloth as J. S. Mill, somehow escaping the German *Methodenstreit* of the 1880s and 1890s, out of which came

the ideal-type. He admits that in saying this he has "gone one step beyond the mandate of the preceding discussion," a very big step indeed. His point of departure is completely unobjectionable and was already established a century ago, that is, the researcher responsive only to "facts" and the procedures of their capture is not doing science. But from this article of faith he struggles to reach a justification of his own work: "If empirical research alone will not invalidate more general assumptions, if the result is rather to produce a more elaborate generalizing strategy, and if this theoretical struggle is necessarily a decisive element in any major shift of social scientific commitment, then it becomes a matter of the greatest importance to make explicit the logic upon which such general debate depends" (p. 35). At the highest reaches of the Vatican, among its most learned keepers of the dogma, there might be heard an argument as hypothetically and idealistically arranged, and as distant from daily Catholicism as this is remote from daily social science. But out in Montaillou, the parishioners have always known the score. If the Vatican wants to employ bright men to argue fine points, what bearing does this have on "Catholic community commitment"? If epistemologists and logicians feel the ground shift beneath them every time they encounter the irrational, will social research pause in hopes they can repair the damage? No one can fault Alexander's goal, to provide not so much a "theoretical logic" as a "logic of theoretical change," in the hope that through its application something like consensus might occasionally bless the wayward *Geistes-wissenschaften*. But who could imagine he will succeed? Even Husserl gave up a similar task after 40 years of trying, and that is why his last work is his best. We live in a theoretical Babel, but reducing the cacophony might not increase our knowledge. Instead, it might so becalm the swirling dialectics of paradigm confusion that static conformity to thematic prearrangements would result. Is that not what Galileo encountered when he began restructuring theoretical logic?

But Alexander does not think so. For him "the continuum of social scientific thought" is eminently specifiable, hence a weapon for correcting dissident views and practices, which he applies throughout his work. A table (p. 40) describes the continuum as moving (the two-tiered notion) from the "metaphysical environment of science (nonempirical)" to the "physical, empirical" environment. Between the two lie presuppositions, ideological orientations, models, concepts, definitions, classifications, laws, complex and simple propositions, methodological assumptions, and observational statements. And along the line fall various squabbles, over order and action, ideology, systems, conflict versus equilibrium, and the philosophy of science. We have seen such tables before (many in the 20 years after World War II, the heyday of scientism within the discipline), and there is nothing categorically wrong with gluing together another

one. The question is, Does it say anything new, or, more important, does it illuminate in a fresh way?

Alexander explains that dividing up science in this manner occurred to him after reading Parsons, Smelser, and Stinchcombe (p. 127, n. 1), along with Abner Cohen and Bernard Barber on a "multidimensional theory of action and order" (p. 66, passim) which promises to defeat subject-object dualism. Alexander sets great store by "multidimensionality," both here (pp. 122–26, the text's last section, which is followed by many pages of notes) and in the *ASR* article (1978, pp. 183, 184, passim), but his multidimensional theorizing turns out to be a minor rewrite of Parsons's attempt in 1937 to balance within "a state of tension" the "normative and the conditional" (quoted in Alexander, p. 68). Yet for all this pedigree, I suspect he operated in a literary rather than analytic way. He saw how many debates within the field endangered Parsons's standing, identified which texts he considered major or representative, then arranged them on his "continuum of scientific thought." This is his framework for the volume. Thus he writes on "the reduction of general logic to" political commitment, method, conflict, and functionalism. But he spends nearly half the text pages (pp. 64–112) considering "the generalized problems of action and order," a heuristic program borrowed once again from Parsons. Sorting this out takes Alexander into debates with Mannheim (ideology), M. J. Mulkay and Habermas (positivism), Dahrendorf, Lockwood, and Collins (conflict), and many others on functionalism. It is a brave, or foolish, scholar who takes on Mannheim's little masterpiece, "Conservative Thought," in two pages (pp. 43–44) and then claims that "by pointing to the weakness in the explanatory power of Mannheim's principal categories, one can present, in effect, 'empirical' evidence for their lack of generality and decisiveness" (p. 44). Again Halévy ([1901–4] 1972) is drawn on to supply the "in effect 'empirical' evidence." Mannheim's essay is loaded with precisely that, but presumably the wrong kind.

As suggested by comments made earlier, I cannot agree with many of Alexander's deconstructions. But he might argue that the value of his formal scheme is not vitally tied to the quality of his interpretations. (This is a side benefit to modeling himself on Parsons, who for years defended his first book [(1937) 1968] as not principally a history but a synthesis— which makes pp. 129–694 of that work hard to assess.) What is problematic, though, is accepting his tabular summary of science as better than others, and discovering why he believes that pondering it will lead one to a "multidimensional theory of action and order" of a neo-Parsonian type. This is the great chasm in the book: Alexander's belief that by pulling down all the critics Parsons prompted to write, one more or less indubitably ends up where he did, refurbishing the functionalist mansion.

He joins many others in the fond belief that science is science, ideology is ideology, and *real* scientists know the difference. Only backsliders (like Mannheim, Dilthey, Habermas, Giddens, Friedrichs, Kuhn, Feyerabend, sometimes Weber) would disagree. First and foremost, Alexander is a scientist, which takes us to my next objection.

Despotism of the Concept

All along I have complained that Alexander moves *from* assertions loaded with his special values *to* texts useful in "proving" or illustrating them, positively or negatively. This is antihermeneutics, since the integrity or meaning of the text is seldom given even minimal concern. In effect he reverses the worthy interpretive procedure of moving from a text to an argument. Here is a standard example: "In social theory, the issue of voluntarism revolves, on the most general level, around two long-standing debates, the arguments over nominalism vs. realism, and subjectivism vs. objectivism (Aristotle, 1962: Bk. 7, 3, 4; Plato, 1945: 80–5, 88–92, 321–36; Augustine, 1948: Chap. 14; Halévy, 1901–04; Stark, 1962; Sartre, 1968; Martindale, 1960; Hughes, 1958; Habermas, 1973a; Wilson, 1970). Parsons's position on these questions has been as radically misinterpreted as his position on . . ." (Alexander 1978, pp. 178–79). What is a citation for? Is it pedantry to ask whether the fourteenth chapter of *The City of God* or an unspecified portion of *Search for a Method* concerns these "long-standing debates"? And why is Bierstedt's most famous and germane article, "Nominal and Real Definitions in Sociological Theory" (1959), omitted? Could ideological malice be at work? What is the point of these "dump-bibliographies"? Mere form, decorative foliage, anticipatory defense?

I do not think the writings of precursors, adumbrators, detractors, even allies matter that much to Alexander. Recall his uncharacteristic contradiction of Bourricaud, who saw Parsons as "generous to a fault" regarding his "classical predecessors": "But Parsons was a great thinker, and what Harold Bloom has shown for the great poets [in *The Anxiety of Influence*?] is true for the intellectual innovator as well. Parsons was actually quite ruthless with his predecessors: he meant his work to supersede theirs, not just to codify it. Even when he was brilliant in his insights about Weber, Durkheim, or Freud, these interpretations were revisions, new thoughts launched in response to not only ambiguities in their classical work but Parsons's own future theoretical agenda" (Alexander 1981a, p. 505). The autobiographical element is unmistakable. Alexander has absorbed enough of the "internal normative standards of judgment" within academe to use citations. But he would probably prefer to dispense with all the wrongheadedness around him, and present, like Hegel, apodeictic judgments.

If we scrape away the protective references—which, I maintain, do not bear essentially on his argument anyway—several pivotal claims, expressed as unmediated concepts and relations, become manifest.[7]

Like Parsons, Alexander plays at pleonasm and neologism, even if with somewhat more restraint. Thus it is difficult to present concisely his major contentions in the form in which he lays them out in *Logic* and in the *ASR,* since, as may have been gathered from the quotations, he sacrifices succinctness to the mannerism of tortured qualification. But something else is going on beyond the search for the perfect subtlety. An intense ideological program is at work, tied to what Parsons late in life called "the growth of institutionalized individualism" and to Parsons's acceptance of "the classical liberal emphasis on the autonomy of the concrete individual, although this autonomy is, once again, a multidimensional one" (Alexander 1978, pp. 186, 184). Alexander is not at all shy about admitting this, but he uses an interesting device to distance himself from the unconcealed politics of Parsons's last books in hopes of proving himself the better "scientist." He argues that "the individual" exists in two forms, within both "concrete and analytic frames of reference." The concrete individual is the one reading these lines. The analytic one is less palpable: "But when such empiricism [sensory validation of existence] is penetrated, when this person is viewed analytically rather than concretely, we can see that he is, in fact, a composite of different social forces, the most important of which are the symbolic forces which contain normative elements. Since these elements are internalized, they are in empirical concrete terms invisible; hence, when we look at an individual person, he appears to be discrete when in fact he is interpenetrated with other individuals by virtue of shared symbolic norms" (1978, p. 180). To begin with, how does one "penetrate" empiricism? Here ideology "penetrates" language, since empiricism, mere facts, the stuff of phenomenology, are all ranked below the celestial reach of "the analytic." This is the beginning of what I mean by the despotism of the concept or the conceptual. Not only is Alexander insensitive to language, he seems ignorant of the "linguistic turn" in philosophy and the human sciences. While he may seem to gain, by ignoring this development, a more robust presentation of his ideas, some regard for the theoretically determining quality of language itself, *prior* to conceptualization, could only have helped his argument.

[7] Alexander's four-volume work is to be divided thus: (1) general theoretical problems, (2) Marx and Durkheim, (3) Weber, (4) Parsons (*The Modern Reconstruction of Classical Thought*), the last obviously meant as the apogee of current sociological theory (presumably until Alexander can generate his own). It makes sense here to deal with central arguments not only within vol. 1 of *Theoretical Logic* but also in the *ASR* article, since it seems likely that the latter contains, in condensed form, the main points of the fourth and most important volume. He refers to this longer volume several times in ways that suggest strong continuity between the two works.

But even allowing such "penetration," what does the rest of the passage mean? It seems a return to Simmel more than to Weber or Durkheim. The phenomenal world of human life, if observed momentarily, does not reveal causes for action or inaction, it is true. Why is this not itself, though, an empirical judgment based on observing people behave over time? And though Alexander makes much of Marx and Habermas as proto-Parsonians, stressing as they did—the former more than the latter—the impact of "material" constraints on individual action, he insists that "the most important" social forces "are the symbolic forces which contain normative elements." Throughout his book, and his other works, Alexander shuns idealism and embraces a de-romanticized analytic theory of action (1978, p. 179)—a de-subjectivized theory, if that can be imagined—but he is an idealist nonetheless, as was Parsons. He is at times ingenious in trying to appear otherwise, but he cannot escape his own tail. For me this is not a serious problem, since his and Parsons's political values are mostly laudable from a left-liberal point of view, that is, both in their concern for the individual and in the primacy accorded societal needs. But in trying so hard to dilute the actual presence of the normative in his theory—or balance it off with cosmetic reference to the "material"—he is forced to construct categories without content.

It is not only the "concrete" versus "analytic" individual that leads me to this conclusion, but other things as well, for instance, his belabored distinguishing of "formal" from "substantive voluntarism" or freedom: "If formal voluntarism refers to a universal property of all action abstracted from time and space, and from any specifically ideological properties, substantive voluntarism refers exactly to the opposite: to the degree that particular historical and social conditions allow the realization of individual freedom defined in terms of a particular ideological perspective" (1978, p. 184). He explains that Parsons was ultimately interested in a society that promoted substantive voluntarism for concrete people (perhaps like Cambridge ca. 1950) but felt compelled to spend most of his energy showing that previous writers did not deal enough with formal voluntarism for analytic people—who exist only on the page.[8] Alexander tries to improve on Parsons's approach by claiming that "levels" of importance exist within "scientific thought," and that criticizing theory on the basis of its ideological import is less "theoretically significant" than evaluating its "epistemic presuppositions" (1978, p. 178, n. 1; *Logic,*

---

[8] I suspect I could in turn be accused of "reductionism and conflationary reasoning," because I refuse to be awed by Alexander's banal aperçu that material constraints on life are mediated by "normative interpenetration." Marx refused to be awed by Feuerbach, who had refused to be awed by Hegel for identical reasons. And while neither Marx's nor Feuerbach's readings of Hegel are adequate as hermeneutics, for the purposes of *social* theorists—as opposed to ontologists, like Heidegger—their objections were vital and well taken.

p. 40). He truly sees scientific thinking as representable by the continuum mentioned above, with "presuppositions" coming first, followed by "ideological orientations," "models," "concepts," and so on, marching sequentially from the nonempirical "metaphysical environment" (a mysterious place) to the "physical" and "empirical" environment. This master continuum is Alexander's given, his *obiter dictum*. The canonical statement that epistemic presuppositions (never satisfactorily defined, except by contrast with other writers who failed to appreciate their centrality to theorizing) are "theoretically more significant" than ideological motivations is itself, of course, an ideological claim. This should be obvious even to non-Marxists. The ideology at stake is scientism, in this case working against other ideologies formerly subscribed to by sociologists, like humanism, socialism, liberalism, and so on. But life is ideological; thought is political. To claim otherwise is to forget the first and easiest lessons of sociology, the Durkheimian dicta on the nature of group life. When the theorist pushes those aside—as when he spurns a philosophical anthropology—he begins ascending into the ethereal.

In linking himself resolutely to Parsons, Alexander writes, ". . . it is important to emphasize my belief that one of the most significant and least debatable contributions of that famous discussion [*Structure of Social Action*] is its articulation of the relationship between specifically sociological concerns and the general epistemological issues that have traditionally occupied Western thought" (*Logic*, p. 67). Alexander feels happy and correct in connecting with Aristotle, Augustine, and the rest, and I do not blame him. It is more humanly gratifying, much less wrenching, to deal with the "theoretic-epistemic"—"The logic here is fundamentally theoretic-epistemic rather than ideological or empirical" (1978, p. 182)—than with the calamity that is modern life. It is important to him that "by virtue of its instrumentalist position, Marxian theory is in the same *logical* camp as theories of political Hobbesianism or realpolitik, despite the latter's ideological incompatibility . . ." (1978, p. 182, my emphasis). This is the sort of genuine conflation that Smelser regards as "the most important effort in theoretical synthesis of its generation." Perhaps this needs qualification. If you were worried that Parsons had been forgotten, that functionalism had been defeated, and if you find "logic" more important to social theory than the experienced conditions of existence and their analyses, then perhaps Alexander—who speaks primarily for the generation of the 1950s, *not* the current one—has served you well. As he explains himself:

> Encouraged by the work of Gouldner and Friedrichs as well as by a certain reading of one strand of Habermas' work, this second phase of the sociology-of-knowledge tradition culminates in the current tendency to perceive the

decisions involved in theoretical logic as sufficiently described by the choice of "empirico-analytic" versus "critical" sociology.

The basic rationale for this approach to a postpositivist theoretical logic must be rejected. . . . The ideological dimension plays a fundamental role in every sociological statement, but at the same time the general nonempirical presuppositions that inform sociological reasoning can in no way be reduced to them. In one sense, the remainder of this chapter, and indeed, the rest of this entire work, constitutes an argument in support of this criticism. [Logic, pp. 43–44]

Alexander indulges himself in hunting down at great length "the general nonempirical presuppositions that inform sociological reasoning" and, having done so, claims theory and the discipline have benefited. I am no longer sure what sociological reasoning is, though rereading Mills always helps, particularly after Alexander. But if this kind of "reasoning" is genuinely "nonempirical," why is it sociological? And if sociological, why not ideological? And if ideological, why not liberating? I do not see in *Theoretical Logic* the slightest interest in liberating humankind from its multiple crises. Instead the antique "problem of order" raises its weary head. The high point of humor in Alexander's book is its first line: "If sociology could speak it would say, 'I am tired.' " No; if sociologists could speak together, they would say, "We are frustrated," not because they need to know what Parsons would say if he were alive, but because their ideological and political program—inspired by and creative of 100 years of work—is being ignored by the men who dominate societies they have studied and theorized about. The Reagan administration, for example, would find little to argue with in Alexander's call for revitalization of Parsonian voluntarist theory. In the end, from a political perspective free of agitated qualification, it bespeaks a vision of social life that might have been concocted by a frontier naif who did hard time in German social science seminars. For after the verbiage is "penetrated," the ideology unpacked, what shines through is an unreconstructed liberal American faith in the ultimate goodness of unfettered individual action. This, then, is the fruit of Alexander's extended theoretical intervention put in practical terms. But why does it masquerade as "science"?

Conclusion

Many topics remain untouched, for example, Alexander's terrific difficulties with the problem of irrationality (*Logic*, pp. 72–89); his *deus ex machina*, "multidimensionality" (pp. 122–26); the unresolved, perhaps irresolvable, enigma of "subject/object," something Giddens has also addressed recently; and his disembodied notion of "structure" (1978, p. 183), another Giddens topic. Perhaps these difficulties, and those discussed

above, would have been eased had Alexander pursued some substantive material rather than breathing so much heady theoretical vapor. All the classical theorists began in and never strayed far from substance, allowing their theoretical imaginations to follow what they knew, not precede it vaingloriously. As Charles Page, author of *Fifty Years in the Sociological Enterprise* (1982), has put it: "Superior sociological work has, I believe, four 'dimensions': historical, functional, empirical, and phenomenological. The sociological study of any institutional and cultural complex must reckon with historical context, both internal and external interconnections (manifest and latent), and *both* 'objective' facts and the 'social construction of reality.' This reckoning requires careful attention to diverse kinds of evidence, both logical and sociological reasoning, considerable imagination or playfulness, and analytical skills; it does *not* require obeisance to 'science' " (letter, 1983).

Alexander believes he is involved in "the search for objectivity and rational intellectual standards," and it worries him that Kuhn "has a disquieting penchant for emphasizing" the "vast possibilities for relativism in any nonpositivist position" (*Logic*, p. 113). What if Kuhn and others contributing to "the more radically-antiscientific idealist critiques of positivism" (*Logic*, p. 113) are right? Then social theory must shed its scientific pretense and admit to what it has always been for the Greeks and for the Marxists: the handmaiden of praxis. For example, if social theory cannot help us help others in slipping from the Iron Cage, why bother?

The University of California Press has done a magnificent job with this volume. The dust jacket is tastefully dignified, the typeface is pleasant to read, the ribbon bookmark helps one through the endnotes, which make up half the book, and the paper has a wonderful odor. There are very few typographical errors. Once committed to publishing four volumes of exegeses by a young social theorist, the press has gone all out. Perhaps everyone should buy all four volumes in support of future works the press might choose to publish in social theory.

REFERENCES

Alexander, Jeffrey C. 1976. "Review of T. B. Bottomore's *Sociology as Social Criticism.*" *American Journal of Sociology* 81 (March): 1220–23.
———. 1977. "Review of S. N. Eisenstadt and M. Curelaru, *The Form of Sociology.*" *Contemporary Sociology* 6 (November): 658–61.
———. 1978. "Formal and Substantive Voluntarism in the Work of Talcott Parsons: A Theoretical and Ideological Reinterpretation." *American Sociological Review* 43 (April): 177–98.
———. 1980. "Review of Johan Goudsblom's *Sociology in the Balance.*" *American Journal of Sociology* 85 (March): 1259–62.
———. 1981a. "The French Correction: Revisionism and Followership in the Interpretation of Parsons" (review of François Bourricaud's *L'Individualism institutionnel*). *Contemporary Sociology* 10 (July): 500–505.

———. 1981*b*. "Looking for Theory: 'Facts' and 'Values' as the Intellectual Legacy of the 1970s" (reviews of R. Bernstein's *The Restructuring of Social and Political Theory* and A. Stinchcombe's *Theoretical Methods in Social History*). *Theory and Society* 10 (March): 279–92.

———. 1982. "Three Models of the Relation between Cultural and Social Systems, with Special Reference to the Watergate Crisis in America." Paper prepared for the 10th World Congress of Sociology, Mexico City, August.

Bernstein, Richard. 1978. *The Restructuring of Social and Political Theory*. Philadelphia: University of Pennsylvania Press.

Bierstedt, Robert. 1938*a*. "The Means-End Schema in Sociological Theory." *American Sociological Review* 3 (October): 665–71. Reprinted pp. 31–40 in *Power and Progress*. New York: McGraw-Hill.

———. 1938*b*. "Review of *Structure of Social Action*." *Saturday Review of Literature*, March 12, pp. 18–19.

———. 1959. "Nominal and Real Definitions in Sociological Theory." Pp. 121–44 in *Symposium on Sociological Theory*, edited by L. Gross. Evanston, Ill.: Row, Peterson.

———. 1981. *American Sociological Theory: A Critical History*. New York: Academic Press.

Gabel, Joseph. 1978. *False Consciousness: An Essay on Reification*. Translated by M. Thompson. New York: Harper & Row.

Halévy, Elie. (1901–4) 1972. *The Growth of Philosophic Radicalism*. Translated by Mary Morris. Reprint. Clifton, N.J.: Kelley.

McNeill, William H. 1980. "Review of S. N. Eisenstadt's *Revolution and the Transformation of Societies*." *American Journal of Sociology* 85 (March): 1241–42.

Martindale, Don. 1960. *The Nature and Types of Sociological Theory*. Boston: Houghton Mifflin.

Mills, C. Wright. 1959. *The Sociological Imagination*. New York: Oxford University Press.

Page, Charles. 1982. *Fifty Years in the Sociological Enterprise*. Amherst: University of Massachusetts Press.

Parsons, Talcott. (1937) 1968. *The Structure of Social Action*. New York: Macmillan.

———. 1976. "Reply to Cohen, Hazelrigg and Pope." *American Sociological Review* 41 (April): 361–65.

Sorokin, Pitirim. 1966. *Sociological Theories of Today*. New York: Harper & Row.

Weber, Max. 1975. *Roscher and Knies: The Logical Problems of Historical Economics*. Translated by Guy Oakes. New York: Free Press.

Book Reviews

Forms of Talk. By Erving Goffman. Philadelphia: University of Pennsylvania Press, 1981. Pp. vi + 335. $20.00 (cloth); $7.95 (paper).

William A. Corsaro
Indiana University

Erving Goffman appears to have two focuses in *Forms of Talk*. The first is in reaction to the kind of theorizing in sociolinguistics that attempts to capture the complexity of discourse by the use of highly formalized models. Goffman does not reject formalistic analyses but argues that they should be part of an approach that places talk in a broader interactional framework. A second and related focus is one that recurs in all Goffman's work; that is the demonstration that social interaction constantly involves self-presentation and maintenance. Goffman instructs the student of discourse to look beyond the idealized model of the speaker/hearer as a transmitter of information to the framing devices and strategies through which self is displayed, maintained, validated, or denied.

Toward these ends, Goffman skillfully exhibits the interplay of ritualization, participant framework, and "embedding" in face-to-face communication. In "Replies and Responses," the first of the five essays in this book, Goffman adds to what he calls the "system constraints" of the conversational analysts (e.g., the Sacks-Schegloff-Jefferson "turn-taking procedures") by introducing "ritual constraints." Ritual constraints govern how each individual should handle himself so as not to discredit his own or another's tacit claims to good character. A second important contribution in "Replies and Responses" is Goffman's discussion of discourse units and sequencing. Here his notion of embedding is most relevant. In general terms, embedding is part of our "linguistic ability to speak of events at any remove in time and space from the situated present" (p. 3). Because of this embedding ability, social actors have wide dramatic liberties. We can mimic, mime, reenact, and hide ourselves away from what we have said, are saying, or are about to say. Given this emphasis, Goffman offers a basic interactional unit composed of three moves: mentionable event, mention, and comment on mention. But Goffman does not mean to imply that discourse is simply a matter of chaining a series of interactional moves together. In fact, he argues that responses must have "references" and that "our basic model of talk perhaps ought not to be dialogic couplets and their chaining, but rather a sequence of response moves with each in the series carving out its own reference" (p. 52).

In "Response Cries" Goffman displays his fascinating talent for taking subtle, taken-for-granted features of everyday interaction and showing that these phenomena may not be what they seem to be and that they are much more important than we assume. Response cries are exclamatory interjections (e.g., "Oops!" "Whoops!" etc.) which, along with other types of self-talk and deprecations, are commonly seen as being purely expressive blurtings that often occur at times of stress. Goffman challenges this view and suggests that we "look to the light these ventings provide, not to the heat they dispel" (p. 120). For Goffman, these vocalizations "make a claim upon the attention of everyone in the social situation, a claim that our inner concerns should be theirs, too, but unlike the claim made by talk, ours here is only for a limited period of time" (p. 121). For example, consider response cries that serve as transition displays. These are uttered upon entering or leaving a state of marked discomfort: "Brr!" is the usual exclamation on leaving outside cold for inside warmth and "Ahh!" or "Phew!" the one used when entering a cool place from a hot one. Goffman's point is that these expressions are conventionalized as to form, occasion, and social function and experienced so commonly that they should be studied for what they tell us about social order in everyday life. But Goffman wishes to go a step further. He argues that ritualized versions of these expressions can themselves be embedded in standard conversational encounters. For example, "When a speaker finds he has skated rather close to the edge of discretion or tact, he may give belated recognition to where his words have gone, making a halt by uttering a plaintive *Oops!*, meant to evoke the image of someone who has need of this particular cry, the whole enactment having an unserious, openly theatrical character" (p. 117). In appreciating that these ritualized response cries become part of conversation, one will discover that they cannot be analyzed without references to their original functions outside conversation. Having taken the argument to this point, Goffman ends the essay by recommending "that linguists have reason to broaden their net, reason to bring in uttering that is not talking, reason to deal with social situations, not merely with jointly sustained talk" (p. 122).

For Goffman, a change in "footing" in discourse (the topic of the third essay) "implies a change in the alignment we take up to ourselves and the others present as expressed in the way we manage the production or reception of an utterance" (p. 128). The notion of changes in footing is offered as an alternative to the traditional categories of speaker and hearer. For Goffman, the notion of hearer is too global because it refers to a wide and undifferentiated range of participants. Therefore, Goffman introduces the notions of "participation status" (i.e., one's position regarding particular utterances) and "participation framework" (i.e., one's position regarding all other persons present during discourse). When discussing the term "speaker," Goffman argues that a more useful concept would be "production format." This notion refers to the multiple ways speakers can present themselves (e.g., as animators, authors, principals, or some com-

bination thereof). In "Footing," as in other essays in the book, Goffman ends by demonstrating that the topic at hand is even more complex than he has led us to believe. For changes in footing should not be seen as merely switching from one stance to another. Rather, because of embedding and ritualization, when we change footing "we are not so much terminating the prior alignment as holding it in abeyance with the understanding that it will almost immediately be reengaged" (p. 155). Therefore, one can hold the same footing across several turns at talk, and one alignment can be fully enclosed within another. "In truth, in talk it seems routine that, while firmly standing on two feet, we jump up and down on another" (p. 155).

In the last two essays Goffman applies many of the concepts he introduced in the first three papers. In "The Lecture," he analyzes this form of talk as a speech event and focuses specifically on the comparison of changes in footing in lectures with those of talk in other social contexts. Even though I have given and attended numerous lectures, I had trouble following his argument in this essay. Goffman does make several interesting points, but it is possible that this paper was more successful in its original presentation as a lecture than it is in print.

"Radio Talk" is an insightful and entertaining essay. Goffman begins by describing the special conditions of radio announcing that make speech errors or "fautables" highly noticeable and in need of self-correction. He then goes on to provide numerous examples of various types of fautables and announcers' strategies for dealing with them. An important feature of error correction in radio talk is that in attending to a fautable the announcer directs more attention to it and may make things worse:

> Newscaster: "This is your eleven o'clock newscaster bringing you on the pot report . . . I mean on the spot retort . . . I mean on the tot resort . . . Oh, well let's just skip it!" [P. 310]

Goffman's main point, however, is not just to provide us with amusing examples. He argues that an examination of radio talk can direct our attention to critical features of everyday informal talk that might easily go unnoticed. Informal talk, unlike radio talk, is highly flexible and allows speakers a considerable margin of error. In informal talk the speaker can from moment to moment during discourse "meet whatever occurs by sustaining or changing footing." And most important, he can select "that footing which provides him the least self-threatening position" (p. 325).

In sum, *Forms of Talk* is an interesting and insightful book. I recommend it highly to all of those who are interested in the relationship between language and social life and even more to those sociologists who are familiar with Goffman's dramaturgical approach but unaware of his recent important contributions to sociolinguistic theory.

The View from Goffman. Edited by Jason Ditton. New York: St. Martin's Press, 1980. Pp. 289. $25.00.

Ulf Hannerz
University of Stockholm

Erving Goffman was one of the few individuals in contemporary sociology who have been genuinely innovative theoretical thinkers and who have, at the same time, had an intellectual impact outside their own discipline, or even outside the academic community. Yet there have been varied views as to the real nature of his theoretical stance and the worth of his contribution. Was Goffman a symbolic interactionist or a structuralist? Was he a pop sociologist with the footnotes left in, or a mid-century Simmel? From what class perspective did he really look at society? And so forth.

In *The View from Goffman* 11 scholars have undertaken a searching analysis of Goffman's work. Six are Americans, five (including the editor) British—evidence of a growing interest in Goffman in Great Britain since the 1960s. While all the Americans have published work on Goffman before, most of the British have not.

On the whole, the contributors are appreciative but not uncritical of Goffman's work. Gouldner's discussion of Goffman in *The Coming Crisis of Western Sociology* is the point of departure for two who take positions opposing Gouldner's. Mary Rogers insists that Goffman had more to say on issues of power and inequality than he has been given credit for, while George Gonos proposes that the underlying class position of Goffman's social theory is not that of Gouldner's "new middle class." Gonos notes that Goffman's early writings come out of the same postwar period as those works of Mills, Riesman, and Whyte that are concerned with this expanding class. Goffman's own work, he suggests, takes the position of the lumpen-bourgeoisie, the remaining independent small entrepreneurs, who struggle to find their way through (or around) this new class structure. This class perspective has much in common with that of the lumpen-proletariat that is often the immediate source of Goffman's ethnography.

Gonos also describes Goffman as "an American structuralist." The reaction against the routine assumption that he represented only a somewhat odd variety of symbolic interactionism is recurrent in several of the contributions to this volume. Clearly Goffman's thought must be viewed against a wider background than that of homegrown American sociology. In analyzing Goffman's early writings, John Lofland emphasizes the strand of existentialism, while Randall Collins traces an intellectual network of considerable range. He notes the symbolic interactionism of Goffman's Chicago years but also the presence in Chicago of Durkheimian, social anthropological influences mediated by Lloyd Warner—influences which had already been present in Goffman's earlier academic milieu in Canada. Collins also mentions the ties to Simmel and Kenneth Burke, pays at-

(

tention to the connections with game theory (Herbert Simon and Thomas Schelling) of *Strategic Interaction,* and the influence of the ethnomethodologists and linguistic philosophers on *Frame Analysis.* Collins's conclusion is that Goffman consistently kept in close touch both with the frontiers of academic social thought and with evolving popular culture, perhaps using each to establish his role distance from the other.

The question of whether Goffman is really a structuralist has been raised mostly since the publication of *Frame Analysis.* Here Gonos brings it up again, pointing to an emphasis on rules underlying interaction. Steve Crook and Laurie Taylor, in their contribution, are critical of an earlier Gonos publication with this view, however, and consider Goffman systematically ambiguous on this point, with both interactionist and structuralist tendencies included in his "frame" concept. Peter Manning also touches on this controversy in his elegant concluding essay on Goffman's style and ends by taking a kind of structuralist view.

In other contributions, George Psathas examines *Strategic Interaction* closely and states his own phenomenological alternative; Mike Hepworth shows the usefulness of Goffman's thought in studies of deviance and control but does not attempt to develop it much further; and Robin Williams examines Goffman's conceptualization of conversational interaction, noting its emphasis on accommodation.

While this volume provides a good overview of Goffman's work and responses to it, his writings are not covered evenly. Not surprisingly, *The Presentation of Self* and *Frame Analysis* get the most attention. There is also some inevitable overlap with previous commentary, although contributors on the whole make little direct reference to writings preceding theirs (apart from drawing on some of their own earlier publications).

For a reader who wants to trace Goffman's writings as well as earlier comment on them, Jason Ditton's editorial conscientiousness makes this a particularly useful volume. His introductory "bibliographic exegesis" lists all Goffman's available writings, beginning with the unpublished M.A. thesis of 1949 and continuing up to 1979. It also includes reviews and other published commentary, although in this respect it does not claim to be complete. With one exception (Boltanski), it covers English-language work only. Not included is what is probably the first book on Goffman, published in Danish in 1975 and edited by Gregersen. More surprisingly, Bennett Berger's 1973 *Encounter* piece is also omitted. Because contributors to the volume cite different American and British editions of Goffman's books, the editor has adopted a rather cumbersome procedure for collating references. It would have been simpler to refer to only one set of editions, but possibly this would have displeased either American or British readers.

A book like this does much to illuminate the uses, realized or potential, of Goffman's work and intellectual perspective. Goffman himself was reluctant to respond to his commentators, critics, and admirers, except perhaps obliquely through his further works. So perhaps we will go on, unconstrained by what Goffman may have thought right or wrong inter-

pretations, to find rather different kinds of new problems and inspirations in his writings. One might feel, for example, that more could be said about their relevance for comparative, macrosociological, and historical studies. One anthropologist has taken the perspective of impression management into a study of a Brazilian Indian village. Is it as much at home there as in Chicago? Someone has suggested that Pierre Bourdieu is "a French Goffman"—what is there to such a comparison? The civilizational studies of Norbert Elias deal, on a much larger scale in time and space, with ideas that appear related to Goffman's. Would it be useful to look in greater detail at how their works speak to each other? If *The View from Goffman* crosses some boundaries in starting an Atlantic dialog, its topics might not yet be exhausted.

Essays in Trespassing: Economics to Politics and Beyond. By Albert O. Hirschman. New York: Cambridge University Press, 1981. Pp. viii + 310. $29.50 (cloth); $12.95 (paper).

Shifting Involvements: Private Interest and Public Action. By Albert O. Hirschman. Princeton, N.J.: Princeton University Press, 1982. Pp. x + 138. $14.50 (cloth); $4.95 (paper).

Jan Smith
Ohio Wesleyan University

At the beginning of a recent essay on economic development, Albert Hirschman agrees with Kuhn's contrast between the natural and the social sciences. According to Kuhn, normal natural science—the settled verification, application, and extension of widely accepted paradigms—is interrupted only infrequently by the turmoil of paradigm change. But any would-be paradigm in the social sciences is likely to elicit a barrage of criticism so compelling that nothing gets settled. Luckily for Hirschman, Kuhn is right, for in "normal science" there is no room for Hirschman's craft. He cannot be confined by any paradigm, or looser school of thought, or even by his home discipline, economics. Broad categories—economic development, political conflict, human motivation, the history of ideas—may capture much of his work, but they do not convey its distinctive quality. Hirschman exploits the anarchy of the social sciences, traversing paradigms, disciplines, nations, and centuries to discover congruities and complements among previously isolated achievements.

The title *Essays in Trespassing* acknowledges the author's refusal to stay at home. The 14 essays are divided into five groups, each corresponding to one of his earlier books. But it is the introductory essay, set apart from the others, that offers the best clues to the coherence of his

work. This essay is a retrospective consideration of the character and fate of his first resting-place, development economics. Way back before Simon and Lindblom and the resurgence of Marxism and stagflation, says Hirschman, development economics rejected the orthodox claim that traditional economic theory is valid universally. This rejection must have attracted him then, because it excites him even now. Meanwhile, he has rejected consistently two features of traditional theory: its failure to comprehend politics and its psychological premises. The fruits of his dissent are nowhere more evident than in his essay on inflation. He skips over traditional theories of inflation—wage push, monetarist, and so on—to dwell on a deeper cause, distributive conflict in the political arena. Because he knows Latin American history, he does not stop at the fashionable but naive hypothesis that such conflict is rooted somehow in electoral democracy. Instead, he asks how a distributive conflict made futile by inflation can persist, finds the orthodox Prisoner's Dilemma answer unsatisfactory, and ventures into a psychology that is foreign to economics in order to construct a better answer.

Development economics was not an unrelenting rebel. It accepted the claim of traditional theory that economic relations between industrial and developing societies are mutually beneficial. Hirschman recalls the optimism engendered by this claim, the sense that development economists could provide intellectual aid to poor societies without ruffling the rich ones, but he no longer shares it. He recognizes that developing societies have gained by trade and growing GNPs, but he also sees unbalanced growth, increasing inequality, and the rise of authoritarian states. This is just one instance of his tolerance for a complex compromise between extreme and simple views. *Essays in Trespassing* combines a tolerance for compromise with a talent for synthesis. Hirschman's justly famous synthesis of economics and political science, exit and voice, is elaborated in four essays that range from fission of Nambikuara bands to integration of European states. Perhaps the most timely synthesis in the book is one intended to improve our understanding of Latin American authoritarianism. Hirschman recounts hypotheses that attribute the new authoritarian regimes to particular economic problems and finds that none of these hypotheses can account for all of the regimes, but then, instead of dismissing the hypotheses, he subsumes them under a more abstract one—that frustration over failure to solve *perceived* economic problems provokes an authoritarian reaction. The emphasis on perception signifies a further synthesis of theories that emphasize real economic problems and theories that emphasize the ideological fashioning of those problems. Hirschman holds that frustration is more likely when new and difficult economic tasks are openly and loudly proposed before old ones have been completed.

The introductory essay's third clue to Hirschman is in its mixture of aloofness from social science communities and involvement in the larger society. It chronicles the decline of development economics without lament but allows dismay at the complexity of problems faced by developing societies. Probably it is Hirschman's wanderlust that has detached him from the common obsession with a profession. From his detachment comes a fine awareness of intellectual history's ironies. He makes much of an "alliance" between traditional economics and neo-Marxism against development economics. He suggests that Latin American economists who publicize their societies' economic problems are part of their societies' political problems. He shows that Hegel invented a theory of capitalist imperialism that was ignored by Marx and then reinvented by Hobson and Luxemburg. However, his detachment does not extend to society. He does not write about the problems of developing societies merely to advance theory, and when he does advance theory, as in the essay on staple production, linkages, and development, he never strays far from real social problems. In the last essay, he ponders the separation of social science from moral discourse and concludes that a moral social science is his "dream for a 'social science for our grandchildren' " (p. 306).

Shifting Involvements is the product of one of Hirschman's periodic shifts from the study of developing societies to the study of advanced capitalist societies. The aim of this long essay is to explain why such societies cycle repeatedly through phases of private consumption and public action. Much of the explanation rests on the hypothesis that experiences often fall short of prior expectations, yielding disappointments. Disappointments with private consumption accumulate until people shift to public action where new kinds of disappointments finally push them back into the private phase. The cycle recurs indefinitely.

This argument confronts two difficulties. Notice, first, the asymmetry between private *consumption* and public *action*. Hirschman focuses on action in the public phase because the disappointments that he locates there stem directly from public action which absorbs more time than expected and is unexpectedly limited by the one-vote-per-person rule. He focuses on consumption in the private phase so that he can include disappointments with the consumption of both private durable goods and such *public* services as schooling and health care. But this occludes the boundary between public and private. To which phase does agitation for more public schooling belong? If it is public, then does its success imply there will be an automatic shift to "private" consumption of the additional public schooling?

Suppose that Hirschman had removed this ambiguity by adhering to the conventional distinction between private action-and-consumption and public action-and-consumption. He might then have noticed the probable

similarities between public and private disappointments. Just as persons may be disappointed by the unexpected demands or limitations of public action, so they may experience analogous disappointments in the private realm, especially in jobs and careers. Just as they may be disappointed by the fleeting pleasures of such durable private goods as washing machines and cars, so they may be disappointed by such durable public goods as water mains and highways. But these similarities make the argument that there is a public-private cycle less convincing. Why would persons shift radically and recurrently from public to private and vice versa if to do so were to replace one set of disappointments with a similar set?

Grant that people do oscillate between public and private. There is a further difficulty in showing that most members of a society are at the same point in their respective cycles at the same time. Are individual cycles synchronized into societal cycles? Apart from the reiteration of a popular stereotype—private 1950s, public 1960s, private 1970s—there is little direct evidence for societal cycles in this essay. Hirschman seeks to establish their existence indirectly by identifying two factors that synchronize individual cycles. One is an exogenous event that pulls everyone into the public phase at once. But synchronization of this sort is exceptional; even the gravity of war exerts a highly varied pull on various persons. The other factor is introduction of a new good or service for mass consumption. But new goods and services are rarely introduced to everyone, or even to a majority, at the same time. Initial consumption and subsequent disappointment are spread over years and decades. New goods and services may do as much to disperse individual cycles as to synchronize them.

Shifting Involvements has the flowing style and flood of novel ideas that we have come to expect from Hirschman, but its major thesis is uncharacteristically strained. *Essays in Trespassing* ought to be read first because it has broader scope and stronger arguments.

Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context. By Howard Schuman and Stanley Presser. New York: Academic Press, 1981. Pp. xii + 370. $29.50.

Tom W. Smith
National Opinion Research Center

Recently the Panel on Survey Measures of Subjective Phenomena of the National Academy of Sciences issued its summary report on survey research. One of its strongest recommendations was an improvement in

survey methodology. The panel urged survey researchers to "take surveys seriously in much the same way as physicists take particle accelerators seriously, astronomers take telescopes seriously, and space scientists take space vehicles seriously" (*Surveys of Subjective Phenomena: Summary Report* [Washington, D.C.: National Academy Press, 1981]; p. 42). The work of Howard Schuman (who served on the NAS panel) and Stanley Presser epitomizes the type of survey methodological research advocated by the panel. Their book, *Questions and Answers in Attitude Surveys,* systematically examines many of the major concerns of questionnaire design: question and response order, open versus closed questions, filtered and unfiltered item nonresponse, middle positions, balance, acquiescence, multidimensionality, and question wording. Skillfully using an impressive series of split ballot and test/retest experiments, they seek to uncover the magnitude and form of the measurement variation associated with each of these formats and approaches. While carefully inspecting the differences in distributions, Schuman and Presser also study higher-order effects in order both to examine the hypothesis of form-resistant correlations and to illuminate the cognitive and social psychological factors behind the measurement variation. One of the particular strengths of Schuman and Presser's research is their conceptualization of response effects. Rejecting the telescope model that sees measurement error as annoying imperfections (caused by imperfectly ground lenses or imprecise gauges), Schuman and Presser realize that the measurement variation caused by the various questionnaire forms represents real social and psychological phenomena. Studying these response effects significantly advances our understanding of psychological and cognitive processes and the sociological interaction involved in the interview situation.

Summarizing the findings would be unfair in a sense, because one of their major findings is that there is great complexity in the survey art, that simple rules and facile generalizations do not usually apply. For example, they demonstrate that the common assumption that measurement variation is concentrated among the less educated is usually erroneous. Sometimes, as in some counterarguments, this does occur, but in such other situations as response order or item nonresponse filters no associations appear. Although they cannot provide simple universal answers, Schuman and Presser do increase our knowledge of response effects vastly. For example, their study of question order effects shows that order effects occur among related questions, that general questions are more susceptible to such effects than specific questions, and that questions that imply the norm of evenhandedness create large, consistent, and stable order effects. They offer additionally a typology for understanding the factors that induce order effects. Similarly, their work with measures of importance, salience, centrality, and committed action clearly shows that

much variation escapes when only affective attitude questions·are employed. A two-to-one majority favors gun control, but among those who have written letters or made donations a clear majority opposes gun control. Their work also shows the advantage of a multidimensional approach that couples affect questions with items on such other dimensions as importance, action, salience, and knowledge. Finally, their study of open and closed questions illustrates the practical recommendations that emerge. They suggest that open-ended questions be used in pilot questionnaires in order to devise suitable closed-ended questions for·the final survey.

The limitations of the book are twofold. Response effects are a complex area undergoing pioneering exploration. Many effects are not well understood, and some of Schuman and Presser's attempts at explanation will undoubtedly be overturned by later research. Indeed, they, themselves, were able to overturn several preliminary findings through their rigorous program of replication. This, of course, is true on any frontier of knowledge. Second, Schuman and Presser concentrate almost entirely on simple question forms (dichotomies, trichotomies, and a few Likert scale items). They neglect measurement instruments at the leading fringe of survey methodology, such as magnitude measurement scales and feeling thermometers. By focusing on the types of questions that are most frequently employed in survey research, they provide maximum information about the instruments being used but do not inform us whether better instruments are available or whether their findings on such matters as order effects and crystallization would also apply on these types of instruments. Once we learn how "simple," standard questions work, we should extend our focus to these newer forms.

In summary, *Questions and Answers in Attitude Surveys* makes an enormous contribution to the field of survey methodology and ranks with such classics as Hyman's *Interviewing* and Kish's *Survey Sampling.* To the extent that survey research is only as reliable as its methodology, this book becomes a long-missing foundation stone for the social sciences in general.

To Dwell among Friends: Personal Networks in Town and City. By Claude S. Fischer. Chicago: University of Chicago Press, 1982. Pp. xii + 451. $35.00 (cloth); $10.95 (paper).

Charles Kadushin
Graduate Center, City University of New York

In a classic expository review of *The American Soldier,* Paul F. Lazarsfeld noted a number of "obvious" findings none of which was true (*Public*

Opinion Quarterly 13 [Fall 1949]: 380). It continues to be "obvious" to both the general American public and to Western intellectuals generally, as Fischer points out, that "modern society has disrupted people's natural relations to one another, loosened individuals' commitments to kin and neighbors, and substituted shallow encounters with passing acquaintances. . . . Community has disintegrated into a mass of atomistic and alienated individuals. The moral order, once undergirded by strong links among people, is now precariously maintained by institutions of mass control, by false community" (p. 1). The city is a major villain in the passing of traditional society. "There is a fundamental and useful, if not completely accurate, analogy: city is to small town as modern society is to traditional society, in all the respects just mentioned" (p. 2).

What is "obvious" is once more not true. In *To Dwell among Friends,* Claude Fischer demolishes most of what everyone always "knew" about the effects of urbanization on social and moral life. More important, he specifies conditions under which some myths hold while others do not. Although American intellectuals, frightened as they are of figures and numbers, will probably continue in their errant views, it will be without excuse. Fischer writes clearly and sometimes elegantly, condensing complex strings of multivariate analyses into a few coherent sentences.

The data were collected from 1,050 adult English speakers living in 50 localities in northern California (defined as San Francisco–Oakland SMSA and counties 200 miles north and east of the area). Excluded were places with populations under 2,500, predominantly black localities, and such areas with large institutional populations as military bases and colleges. The sample was further clustered contextually into neighborhoods and "micro-neighborhoods."

In the most interesting part of the design, a segment of each respondent's personal network, or what Rossi has called the "interpersonal environment," was elicited. Each respondent was given an opportunity to name other persons with whom he or she engaged in various exchanges ranging from talking about hobbies and looking after the residence while the respondent was away to borrowing money. Each respondent named an average of about 19 people about whom he was asked for some basic information. Additional information, including the sociometry of who knew whom, was solicited from the respondent for a smaller subset of five or fewer named others.

There are limitations and virtues to the design. Fischer is well aware of the most glaring limitation: it is "obvious" that northern California is unique. Fischer proves to his satisfaction that it is not, but hot-tub hedonism and an integration of rural and urban spaces through recreation remain my view of California middle-class networking. The exclusion of black neighborhoods, while understandable, loses a critical feature of

central cities. On the other hand, the contextual design of neighborhoods and communities which is a major virtue of the study is, for some reason, not exploited in this analysis.

The basic plan of analysis is to take some aggregate figure or index derived from the network data, for example, the proportion of persons named who are kin, attach this to data on the respondents themselves, and then see whether an urbanism index grouped into four points—semirural, town, metropolitan, and regional core—is related to any differences on the network variable. Since there are considerable demographic differences among places with different degrees of urbanism, appropriate statistical controls are then introduced that use the SPSS package to determine whether observed differences caused by the urbanism index can be reduced or eliminated. The network variables are also used as independent variables predicting, for example, psychological well-being. As is typical in surveys, when data on social structure are used to predict individual differences, the amount of variance accounted for is often quite small. Fischer argues well for the practical and theoretical importance of structural variables even when they seem to account for little variance in survey data. The statistical analyses are quite properly used heuristically as guides to interpretation of the data rather than as rigorous tests of hypotheses.

I take these details from the appendixes and the footnotes and discuss them first because I believe my audience for this review are "experts." On the other hand, Fischer is absolutely correct in keeping the analytical apparatus out of the way, yet making it available. If more quantitative sociologists would publish this way, the profession just might get a better hearing.

The findings, in brief: basically there are *no* differences in the amount or quality of social ties between more and less urban respondents, although small-town residents are involved more with kin and with "traditional" relations, and more of them in their interpersonal environment know each other. Nevertheless, urbanism is *not* associated with less social support—if anything, it is the converse. Further, "City and town respondents did not differ psychologically, and we could not find much evidence here (or in other research) that urban life directly or through its effects on networks, seriously affected anxiety or mental health" (p. 260). Yet city dwellers do worry more about crime, and this leads to distrust of strangers. Urbanism encourages social relationships among people who are strongly interested in a particular subject or activity, especially if they are a minority, and, partly as a result, big-city dwellers are less traditional and more receptive to "novel values." At the same time, urbanism decreases relationships based on formal ties such as organizational and church memberships. Fischer concludes that "urbanism tends to produce

232

a different *style* of life, but not a different *quality* of life" (p. 260, italics in original).

I have some quibbles: (1) Fischer recognizes that, in controlling for demographic differences between town and city, he is being conservative; in fact, the differing demographic structures may not only represent selective migration but also be an inherent part of the character of cities or small towns. (2) Based on the data reported, it seems that the networks of persons of lower social class, especially in more urban areas, differ sharply from those of higher-class respondents. I would have made this much more central in the book. (3) The discussion and data analysis of urban subcultures might have benefited by the introduction of Simmel's structural concept of overlapping circles. (4) The network field is itself a rather tight network; a more explicit acknowledgment of previous and concurrent work, both methodological and substantive, in the study of urban networks might have been useful for those who are not "insiders."

None of these reservations is intended to detract from my evaluation of this book as a major, seminal, well-written, long-overdue confrontation with data of myths about the social consequences of the urban environment.

The Growth Dilemma: Residents' Views and Local Population Change in the United States. By Mark Baldassare. Berkeley and Los Angeles: University of California Press, 1981. Pp. ix + 175. $20.00.

Margaret Mooney Marini
Battelle Human Affairs Research Centers

The Growth Dilemma examines the relationship between change in the population size of local areas and residents' attitudes. It focuses specifically on the relationship between change in population size and residents' perceptions of change, and feelings of satisfaction. The book is short. It consists of six chapters and only four of these contain empirical results. In the first chapter, Mark Baldassare presents vignettes of recent patterns of population change in the United States, contrasts these patterns of change with population changes at earlier times and in other countries, examines sources of ideas about population change (including public opinion, planners' and policymakers' perspectives, and sociological theory), and outlines several methodological approaches to the study of population change.

Chapters 2–5 contain the results of empirical analyses based primarily on two data sources: (1) Cycles VI and VII of the NORC Continuous National Survey, which consists of household interviews with 886 met-

ropolitan residents and was conducted nationwide between 1973 and 1974; and (2) the Quality of Life Survey conducted by the Institute for Social Research of the University of Michigan in 1971, which consists of interviews with a national sample of 2,164 individuals. To each respondent's record in these data sets were added measures of population change in the local area based on census reports of the population and housing characteristics of the respondent's census tract, county, and metropolitan area. Most of the analysis of change in population size focuses on a measure of countywide net migration between 1960 and 1970.

Chapter 2 examines the relationship between actual countywide population change and respondents' perceptions of change, its impacts, and preferred interventions. Chapters 3–5 examine the relationship between actual population change and residents' feelings of satisfaction with (1) local services and facilities, (2) physical and social aspects of the residential environment, and (3) social relationships and personal well-being. Each chapter focuses on residents in a different type of community. Chapter 3 focuses on residents of central cities; chapter 4, on suburban residents; and chapter 5, on residents of nonmetropolitan areas. In the last chapter, the author summarizes the major findings of earlier chapters, discusses implications for future research, and makes predictions about future population change.

Although the implications of change in population size for residents' perceptions of change and satisfaction have not been studied previously, the contribution of this book to our understanding of these issues is limited. Much of the book is a presentation of well-known information, available from the U.S. Census, about changes in the population size of different types of communities. This descriptive information not only is presented in the introductory chapter; each chapter containing empirical analyses has a lengthy introduction. The introductory chapter also contains a discussion of theories with little relevance to current population shifts. These propositions, purported to guide the empirical analyses in later chapters, have no obvious origin and appear to have been derived post hoc.

The empirical analyses are impaired by methodological weaknesses. The measure of population change used poses a critical problem. A *countywide* measure of net migration over a decade is used to draw inferences about population change in areas *within* counties. Since a county may include several types of communities (e.g., a city, suburbs, nonmetropolitan areas), and parallel changes in the sizes of these communities are unlikely to occur, the net change in population size for a county is unlikely to be indicative of the population changes in individual communities. Within a county, the central area of a city may decline; some suburbs and some nonmetropolitan areas may grow, while others

decline. For only some of these communities is a measure of population change for the county as a whole likely to be even approximately accurate. Although the measure of population change used was considered to be the best of several measures examined, the fact that it bears little relationship to residents' perceptions of change in the size of their communities provides some indication of its inaccuracy.

Most of the analysis examines the effects of net countywide migration on residents' feelings of satisfaction. Previous research suggests that feelings of satisfaction of the type examined are unlikely to be affected much by a variable such as change in population size, since they are affected primarily by immediate personal circumstances over which an individual has some direct influence. Analyzing measures of satisfaction is also problematic because these reflect cognitive judgments made against some standard of comparison, such as aspiration level. To the extent that those in different types of communities have different levels of aspiration, changes in population size are likely to bear different relationships to satisfaction.

Because the statistical methods used are not state-of-the-art, they raise other questions about the accuracy of the findings. Most of the variables are measured on ordinal scales with a small number of categories. Correlation and regression analyses, as well as some bivariate analyses of mean differences, are performed.

In summary, this book attempts to extend our knowledge of the effects of change in population size on the quality of life in a community by focusing directly on residents' perceptions and feelings of satisfaction. Linking a macrolevel phenomenon such as change in population size to outcomes at the microlevel is a difficult task. Successful accomplishment of it requires that the macrolevel measure be refined enough to reflect the actual circumstances that confront individuals. This book contributes little to our understanding of the effects of changes in population size on residents' attitudes because the macrolevel measure of population change used is too crude to reflect the community circumstances experienced by individuals.

Machine Politics in Transition: Party and Community in Chicago. By Thomas M. Guterbock. Chicago: University of Chicago Press, 1980. Pp. xxii + 324. $20.00.

William J. Grimshaw
Illinois Institute of Technology

Despite its being marred by what some readers will consider a fatal flaw (to be discussed later), Thomas Guterbock's *Machine Politics in Tran-*

sition is an impressive piece of research. The theoretical issues addressed are significant, and the research is thoughtful, rich, and meticulously done. The focus of the study is on how the machine secures electoral support, and the author examines one of the Chicago machine's ward organizations to find an answer.

Guterbock begins by constructing three models of machine support. Two of these are conventional: the material-exchange and affectual-exchange explanations. The third, more novel explanation, which he calls the commitment model, involves a normative argument. Electoral support "is won by the party's continuous efforts at legitimizing itself" (p. 10). In order to measure up to the electorate's ideal motives, the machine must foster an image of localistic concern and responsive leadership (pp. 192–206). The theoretical and humanistic significance of the commitment model is that it "does not assume the motivations of machine supporters to be fundamentally different from the motivations of other voters" (p. 9).

The author tests models by studying two problems: how the ward organization secures the compliance of its workers, and how it secures electoral support. Guterbock's findings on the first will surprise many readers. It appears that the machine is poorly equipped to get much more than a modest effort out of its legendary patronage army. The machine does have jobs at its disposal. But salaries are modest, mobility is limited, firings are rare, and performance standards are vague (pp. 35–46). Thus, what the machine must accomplish here is similar to what it must do with the electorate: legitimize itself by measuring up to the ideal motives of the workers, and, given the venal nature of the machine, this is not easy (pp. 67–68).

In its attempt to secure electoral support, the machine is constrained by a similar set of problems. The quantity of the machine's services does not appear to have declined over the years, but the quality certainly has. Those venerable staples, the "fix" and the bonds of intimacy between party agents and voters, are now all but gone and have been replaced by such low-value items as information and referral which are provided by precinct captains, many of whom no longer even reside within the ward (p. 28).

Guterbock's most telling point against using the exchange models comes, however, from a sample survey of ward residents. He found that those who do not make use of the machine's services are actually more loyal to the machine than those who do. The nonusers are the loyalists, the author concludes, because of their low social standing and low political participation rates: "for those voters who lack personal linkages to the local political system, party loyalty functions as a simplifying guide for electoral choice" (p. 169). The users, on the other hand, are much more

236

discriminating. Party services, they believe, are something to which they are entitled by right, not favors to be reciprocated with votes.

Thus, the question arises, If it is not by the provision of services, how does the machine acquire its electoral support? The study's findings are less compelling on this issue. In part, what happened was that Guterbock anticipated that the exchange models would be validated by the survey results and so did not include items that would have helped test the commitment model in the questionnaire (p. 124). Accordingly, the argument for the commitment model uses primarily qualitative data derived from the author's participant observation—he served as a student intern to a ward alderman for two years.

The study's principal deficiency involves the ward and ward organization used to test the models. Although, in the study, it is thinly disguised for reasons of confidentiality, the organization is easily identified as that of the Forty-sixth Ward, one of the weakest and most unusual in the machine. It consistently ranks near the bottom of the city's 50 wards in terms of plurality production. (In the last four mayoral elections, it ranked forty-sixth, forty-fifth, forty-sixth, and was thirty-fourth in the race won by Jane Byrne.) "Amateur" Democratic clubs and even radical organizations are active in the ward. The ward is represented by an "independent" in the state legislature. So, it is not surprising that the ward's machine-slated alderman behaved more like an independent than a regular. This behavior was carried to such extremes that he was dumped by the ward committeeman after two terms (p. 299).

This is decidedly not the stuff that makes up the machine. What effect the unusual nature of the case has on the study's findings is difficult to determine. Guterbock's arguments and research and my own research on the city's black politics convince me that the commitment model is essentially sound. However, we need more evidence.

Informed Consent in Medical Therapy and Research. By Bernard Barber. New Brunswick, N.J.: Rutgers University Press, 1980. Pp. viii + 203. $14.50.

Robert M. Veatch
Kennedy Institute of Ethics, Georgetown University

In *Informed Consent in Medical Therapy and Research,* Bernard Barber adds to his already important contributions to the sociology of medicine. He is perhaps best known in medical sociology for his valuable empirical study of the conduct of medical research on human subjects (*Research on Human Subjects: Problems of Social Control in Medical Experimen-*

tation, coauthored with John J. Lally, Julia Loughlin Makarushka, and Daniel Sullivan [New York: Russell Sage, 1973]). In the present volume he expands the study of the mechanisms of social control in medicine by thoroughly analyzing informed consent from a sociological perspective.

It is perhaps important to understand first what this volume is not. It is not another empirical study, although it summarizes most of the important empirical work on the consent process. It is not primarily a philosophical analysis of informed consent, although the chapter "Values, Norms, and Informed Consent" reveals the links between contemporary philosophical discussion and the sociological themes the author is pursuing. It is not primarily a review of the legal requirements of informed consent, although the chapter "Legal Principles, Legal Rules, and Informed Consent" summarizes the current legal controversy, especially as it bears on important sociological issues on Barber's agenda. It is not participant observation of the sort that Renée Fox and, more recently, Charles Bosk have offered. It is not even primarily a handbook for researchers who need more skill in obtaining an adequately informed and free consent, although this designation is much closer to Barber's objectives. The volume does contain observations about what makes the consent process difficult and some good suggestions for improving it.

The book includes all of these, but most significantly it is a detailed examination of the sociology of informed consent and the social control mechanisms that make the consent process so problematic. Three important chapters in the volume analyze the problems of informed consent in terms of (1) authority and power relations, (2) problems of communication, and (3) socialization and social control of the consent process.

Barber emphasizes that the consent process is one that takes place in a social system involving authority and power. He draws on two classical formulations of the patient/physician relationship: the collegiality model of Talcott Parsons and the dominance model of Eliot Freidson. The former, involving cooperation and minimization of dissonance, is at best an ideal, although that is sometimes overlooked by those obtaining consent for treatment or research. Some would argue that it is unacceptable even as an ideal. The dominance model, Barber concludes, is closer to the reality of medical practice. Whether he sees the dominance model as normative as well is hard to tell. In any case, this conclusion leads Barber toward practical steps to minimize the adverse impact of such authority relationships.

These appear in part in the chapter on communication, which explores the determinants, structures, and processes of the immediate transmission of information. Here competence and rationality are examined as conditions for adequate communication. A particularly insightful discussion

appears under the heading of "perfectionism." By perfectionism Barber means the often heard argument that "perfect communication" with patients is impossible (because they lack either adequate knowledge or time), and, even if it were possible, it would so distress some patients that the physician ought not even strive for it. Barber concludes that the goal is not perfection but better communication, and he concedes that better communication may require that patients be distressed. The chapter closes with a discussion of new patterns of information flow and suggestions for improvement in consent procedures.

The chapter "Socialization, Social Control and Informed Consent," by tracing the socialization of the physician during his or her years as a medical student and then as an intern and resident, shows how attitudes are transmitted to new members of the profession. This chapter also explores the relationship between informal and formal control mechanisms with emphasis on the increasing importance of the more formal mechanisms.

One of the most important themes in the volume is the shift toward greater public accountability for the standards of acceptable informed consent. Barber shows how approval of colleagues and the county medical societies is being replaced by more publicly legitimated standards: institutional review boards mandated by government, state licensing boards with governmental sanction, and the courts. Barber concludes that "even with considerably improved professional self-regulation, it is inevitable that some governmental regulation will persist: to define larger social goals for the professions, to maintain orderly relationships among competing interests and groups, and to be a court of last resort for abuses the professions cannot control themselves" (pp. 189–90).

The logic of the movement to more public accountability is powerful, perhaps even more powerful than some of its defenders realize. This is shown in the development of the "reasonable person standard" for consent discussed by Barber. According to this newer standard, in order for there to be an adequately informed consent, subjects must be told what the reasonable person would want to know or would find material to a decision to consent. This replaces the older standard in which physicians simply had to disclose what their colleagues would have disclosed in a similar situation. The new standard recognizes that, because people are sociologically different, some people may reasonably want information that it has not been customary to disclose.

This means that failure to get an adequate consent may not always be a simple failure to communicate or a purposeful deceit by the physician. It may be a more complicated situation in which the physician acts in good faith, in conformance with the standards of his group, but still fails

to meet the standard of the patient population or certain subgroups within it. This also means that, in principle, groups made up of peers of the clinician or researcher may never be able to determine by their own judgment whether consent information is adequate. Barber's analysis leaves us with a paradox unappreciated by those not sensitive to the sociological dimensions of the consent process. Review of research and therapy by the peers of the clinician may be necessary in order to assure that certain standards of technical adequacy are met, but, by the very meeting of these standards of professional competence, peers can be identified as sociologically unique and therefore unfit for making judgments from the patient's or subject's point of view. Only by increasing awareness of the sorts of sociological determinants of an adequate consent presented to us by Barber can such problems be resolved.

Conceptions of Kinship. By Bernard Farber. New York: Elsevier–North Holland, 1981. Pp. x + 250. $27.50.

Joan Aldous
University of Notre Dame

A book by Bernard Farber is an event worth noting. He is a sociologist with an unusual record as prophet, family sociologist, and researcher. As a rule, sociologists tend to be prophets without honor because their foresight is generally blurred. Not so with Farber. Almost 20 years ago— in 1964, to be exact—he stated the concept of *permanent availability.* According to this, everyone, regardless of marital status, is a member of the pool of marriage eligibles. At a time when the familistic 1950s were still fresh in sociologists' memories, it was a daring idea. The rising divorce rate during the strife-torn 1960s and 1970s, however, confirmed its prescience. Farber is also one of the few researchers whose study of families takes a sociological rather than a social psychological perspective. In recent years, he has been observing how events in the broader society affect families. He is less concerned with the internal dynamics of families existing within their own private worlds. Moreover, his interest is a sustained one in which he has designed and carried out a series of interlinked studies to test the hypotheses of his theory.

Conceptions of Kinship is the result of this intellectual process. The structural variable Farber has been studying for the past decade is family law and its varying models of kinship ties and genealogical distance and inheritance patterns. Continuing his interest in the "interweaving" versus "differentiating" functions of kinship for society, Farber's theory focuses

on the association of these models with universalism or pluralism. Universalistic models emphasize the individual or the interweaving of individuals into society. Pluralism, however, as a principle of collaterality, perpetuates societal differentiation into special-interest groups with specific ethnic, religious, or socioeconomic identities.

The book is a report of how the kinship models conform to the usual ideas of kin ties and their family life and kinship interaction. Genetic code and canon law models are "universalistic" in their individual perspective, while the Parentela orders and standard American models with their lineage stress are "pluralistic." Civil law model adherents fall somewhere in between these two sets of models. The findings are based upon survey data from ever-married residents, 18–45 years old, in Phoenix, Arizona, and on an additional sample of Jewish families.

On the basis of this general theory, one of Farber's hypotheses was that holders of the pluralistic codes that ensure family identity over the generations would be characterized by more religious endogamy, later marital age, fewer working mothers, and less divorce than would the adherents of the more individually oriented codes. Fertility, he argues, should be higher among the descendant-oriented Parentela orders respondents than among the ancestor-oriented standard American respondents. His theory provided no fertility predictions for the other models.

His findings are complicated because they are affected by religious affiliation, social economic status, and generation (Farber asked about parents as well as about the families of procreation of ego and ego's sibs). Generally, the results support the argument for parents and sibs of those respondents categorized as in the Parentela orders and the canon law and genetic models with respect to divorce and religious endogamy with controls for SES and religion. Results on fertility are less clear-cut. Among respondents themselves, the results were mixed. Those in the Parentela orders, regardless of religion, were more apt to be churchgoers and less apt to intermarry, as hypothesized by Farber, although the reverse was not true of respondents in the canon law and genetic model categories. Moreover, a lesser propensity to divorce, when controlled for religion, was not found among the former group, although, as he predicted, those in the latter categories were more apt to divorce. The fertility findings were generally consistent with the hypothesis when controlled for various possible confounding factors.

Given the roughness of Farber's measure of collaterality, the empirical linchpin of his argument, it is a tribute to the strength of his rationale that his data support it as well as they do. Farber's decision as to which of these models his respondents fitted was based on their responses to just one question that asked which relatives should have priority in re-

ceiving property if someone died without leaving a will. The respondents were presented with nine paired comparisons between relatives for whom the models differed in their calculation of genealogical closeness. The criterion of assignment to a model was the respondent's conformance to its pattern of collateral priorities in a majority of the comparisons. One of the problems that appeared immediately was that over 40% of the respondents did not meet this criterion. Examination of their modal responses led Farber to create another model, the standard American model. Because of its residual nature, one important comparison that is relevant to the nature of the model that Farber posits between grandparents and grandchildren is missing. And despite the many tests Farber ran, I retained a nagging doubt about whether the persons categorized as holding the various kinship models held the ideologies the models supposedly reflected. Sample size limitations, small differences among the models, and the criterion for model assignment raise the possibility that other variables like age, religion, or SES better explain the results—despite Farber's attempts to control for the more obvious confounding variables.

The theory, however, is an interesting one, tying together concepts from political and legal as well as kinship scholars. The relation of the kinship models to behavior and attitudes, however, is not always clear. Occasionally, Farber comes close to positing a causal relation, with the models of collaterality influencing or implying behaviors and attitudes (e.g., pp. 155, 161, 195). In other places, he writes of the models' "association" with other variables and other factors affecting them (e.g., pp. 18, 42, 145). If the former is the case, Farber does appear to claim too much for the explanatory power of his models. According to him, the "orientations expressed by a model" are not "ephemeral concerns" varying with family life cycle or interpersonal relations (p. 19). Farber argues that Parentela orders, canon law, and civic law models of kinship priorities developed in the past are still valid. Yet, as we have already seen, the model encompassing the greatest number of persons was a residual one, and the increasing adherence to the Parentela orders model among older persons suggests an age, period, or cohort effect determining rather than determined by the models, as Farber himself notes (p. 145).

Moreover, the theory's failure to take into account sex differences in the importance of kinship collaterality is surprising in the light of its operationalization in terms of property. Even though our kinship system is bilateral, property matters have traditionally been a male concern.

These, then, are some of the concerns I had on reading Farber's progress report on his current thinking. This dazzling attempt to tie a number of kinship issues together is better theoretically than operationally. But *Conceptions of Kinship* is worth reading, if only to see what point Farber has reached in his intellectual pilgrimage.

Human Adaptation and Population Growth. By David S. Kleinman. Montclair, N.J.: Allanheld, Osmun & Co., 1980. Pp. xiii + 281. $23.50.

David F. Sly
Florida State University

Human Adaptation and Population Growth is a "philosophical-political" essay that attacks Malthusian and neo-Malthusian ideas about the role of population size and growth in determining the conditions under which man lives and dies. This is made abundantly clear when David Kleinman begins by arguing, "The conclusions are that population is not now and has rarely, if ever, been an issue relevant to the survival of people, or even to their comfort; that it is tangential to issues of societal development, the unequal distribution of power and wealth and the underlying social, economic and political processes" (pp. 1–2). The Malthusian-neo-Malthusian perspective merely reflects the ethnocentric and class biases of one group and directs our attention away from the real sources of the problems, which are the abuse of power by elites and the exploitative and controlled social, economic, and political structures which they create and/or tolerate.

After having made his personal opinion remarkably clear, the author examines a number of social issues broadly conceived of as causes or consequences of overpopulation and rapid population growth. While the basic ideas in each of these cases are not new, David Kleinman's presentation and development of them are lucid, exciting, well organized, and comprehensive. In nearly every chapter he easily cites literature from anthropology, history, economics, public health, and sociology in a well-integrated and critical manner, arguing that those studies that support the Malthusian model either are methodologically weak or that their researchers found only what they went looking for. Indeed, the chapters focusing on the development of the Malthusian model and the chapters examining fertility behavior and its analysis are essential reading, not just for people interested in population, environment, and ecological issues, but also for those interested in social theory construction and the sociology of knowledge.

It is essential, however, for the reader to bear in mind that Kleinman wants to persuade the reader that population growth is not responsible for the world's social and environmental problems. The general approach he uses to accomplish this is to show the range of variation observable under different sets of environmental and demographic conditions and to argue that these show the adaptive capacity of humans and their social institutions. While this approach is certainly acceptable, the literature cited for it is never subjected to the same critical scrutiny as that sup-

porting the Malthusian view. Similarly, while asserting that ethnocentrism and political and economic biases have dominated the Malthusian perspective, the author seems unaware that these same charges can be leveled against its criticism. Neither this nor the following comment is intended to be negative but rather to define the work's focus and the intensity of the author's pursuit of it.

While Kleinman points out very well the logical and empirical problems of the Malthusian perspective, he offers no real integrated and general alternative to it. What I found particularly frustrating is that, in a number of places, the alternatives he suggests (but never develops) are, within the isolated context in which he offers them, quite naive. For example, in chapter 10, where he examines current fertility declines in nonindustrial societies, he mentions that 14 countries had fertility rates that were below 30 per 1,000 in 1975. After briefly reviewing some factors usually thought to influence fertility declines, Kleinman concludes these countries ". . . vary considerably in per capita income and types of prevailing economies though, in general, their populations are fairly literate and may be characterized as more modern than other developing countries. However, the same could be said of some countries where birth rates were not falling so rapidly . . ." (p. 92). Kleinman suggests that population pressure that has caused fertility to be controlled could be responsible for motivating the fertility decline. This is deduced from the observation that the countries mentioned are relatively small, and many of them are islands. Given the structure of the arguments against the Malthusian approach, he never asks the logical questions. Why should population pressure lead to the use of fertility control? Are there small countries where fertility has not declined?

In short, as long as the reader is aware that Kleinman offers no clearly developed alternative to what he destroys, this is a worthwhile book. On the other hand, the work fails to offer a well-integrated and analytical alternative to the Malthusian perspective.

Internal Migration during Modernization in Late Nineteenth-Century Russia. By Barbara A. Anderson. Princeton, N.J.: Princeton University Press, 1980. Pp. xxv + 222. $18.00.

Henry S. Shryock, Jr.
Georgetown University

Internal Migration during Modernization in Late Nineteenth-Century Russia is a revision of the author's 1974 Princeton dissertation. Barbara

Anderson feels that an understanding of internal migration in late 19th-century Russia has relevance to understanding that phenomenon in contemporary developing countries. During the period discussed the Russian Empire had a well-developed statistical system. Here the author uses the statistics on place of birth in the 1897 census and supplements them with annual registrations of migrants from European to Asiatic Russia collected at key destinations.

The first chapter summarizes the hypotheses to be examined, the last summarizes how they work. Anderson divides migration during modernization into two main types: (1) to urban, industrial centers, and (2) to a sparsely settled frontier area. (What proportion of contemporary countries in the process of modernization have frontiers like those of Asiatic Russia or the American West?)

. Given the presumed necessity to use official statistics for provinces and cities, the analysis is essentially ecological in nature, using in- and out-migration rates, for example, the number born in Province A living elsewhere in Russia in 1897 per 1,000 of the total born in Province A and surviving within Russia in 1897. Rather than using ecological character-istics as mere surrogates for the characteristics of individual migrants, Anderson regards the area of origin as forming the environment in which persons are socialized and in which attitudes, including willingness to migrate, develop. An illiterate living in a province with relatively high literacy is thought to have more and better information about potential destinations and to be more receptive to new ideas than illiterates living elsewhere. The evidence is indirect. J. da Vanzo, on the other hand, found that origin employment rates in the United States affect out-migration but only for the unemployed (*Why Families Move* [Santa Monica, Calif.: Rand, 1976, p. vi]). Elsewhere the author draws conclusions about the migratory behavior of Jews on the basis of rates for provinces that contain high proportions of Jews, a practice that is admittedly paralleled by the use of election statistics for areas to draw conclusions about the voting behavior of various socioeconomic groups.

A rather "parsimonious" set of explanatory variables is used throughout most of the regression analysis. These are: soil fertility level, proportion of males with primary occupation in agriculture who had an auxiliary occupation, percentage of military recruits who were literate, workers in industry per 1,000 of the population (not the population of working age), number of workers in secondary industry per 1,000 of the population, natural increase rate, population growth rate, and population density, as well as a distance-type variable for migration streams. Some of these variables are examined for more than one period prior to 1897, with consideration given to the likely lags in their effects. The multiple and

245

partial regression analysis, use of both stepwise and nonstepwise regression, use of standardized ("normalized") and nonstandardized coefficients, tests of statistical significance, and multicollinearity are all clearly discussed.

Kingsley Davis's modification of demographic transition theory wherein another response to transitional growth is out-migration from the high-growth area (rather than reduced fertility) is tested. Anderson chooses to measure population pressure by the natural increase rate instead of by population density. She found, however, that the origin area's rate of natural increase was not a significantly positive determinant of migration to modernized destinations but relevant only to Asiatic Russia in the later part of the 19th century.

Other important findings were: Within a settled area, migration rates are positively related to the cultural modernization and, to a lesser extent, to the industrialization of the origin, but negatively related to the importance of traditional agriculture. Migration rates to a frontier area are negatively related to cultural modernization of the place of origin. To a lesser extent, they are negatively related to industrial modernization of the place of origin. Migration rates to a frontier area are positively related to the importance of traditional agriculture at the place of origin.

Thus, some findings for migration within a settled area and to a frontier area were in opposition. In both situations, however, migration to a specific destination tends to decrease with increasing difficulty of reaching the destination—partly a function of distance.

Destinations like the Urals or Donbass, or even Odessa and Kiev, that were intermediate in modernity between St. Petersburg and Moscow, on the one hand, and Siberia, on the other, tended to draw from origins that were also intermediate in modernity.

It may seem surprising that literacy was higher and population pressure less in areas of lower soil fertility, but it was in these areas that industrialization had proceeded further. Thus, out-migration rates by provinces were positively correlated with their soil fertility.

The inclusion of definitions for such commonplace terms as mean, standard deviation, crude birthrate, and net migration makes one wonder about the intended readership. The cautious editor who required these would have been better employed in correcting several persistent grammatical errors, such as using a plural verb with two singular nouns connected by "or." Tables are well designed, but the photo-offset reduction and faint print sometimes affect legibility. There is an excellent bibliography.

All in all, the book is a real contribution to our knowledge of internal migration. We can see how the state of the art has improved since E. G. Ravenstein included Russia in his classic 1889 article.

Toward a New Strategy for Development. Compiled by The Rothko Chapel
Colloquium. New York: Pergamon Press, 1979. Pp. xiv + 365. $25.00
(cloth); $9.00 (paper).

Elbaki Hermassi
Tunis University

At the same time that the Rothko Chapel published its Houston collo-
quium papers debating development needs and programs from a world
perspective, G. Arrighi, A. G. Frank, I. Wallerstein, and Samir Amin
were meeting to discuss and confront their respective views on the de-
velopment of capitalism. These views are now published in *Toward a
New Strategy for Development*. The increase in the frequency of these
summit conferences is not to be taken as a sign of vitality and scholarly
collaboration but as a symptom of polarization and crisis.

The economists who have dominated the development field since the
1950s were optimistic about the possibility of growth and "orderly mod-
ernization" in the Third World; it was actually an abounding optimism
which, according to Dudley Seers, Marxists shared with neoclassic liberals
even when they differed with them regarding the mechanisms by which
progress can be achieved. Such ideas as linear development and an em-
phasis on capital investment which could be provided by foreign assis-
tance were used to launch policies, on the assumption that economic
growth would obliterate whatever resistance there was to change and
would prove sufficient to bring about advances in the political, social,
and cultural fields.

These policies did achieve significant growth in terms of GNP per capita
(the late Bill Warren argued this point very strongly on the basis of a
wealth of aggregate data), but the central facts remained the same; there
were increasing inequality, growing repression, and widening interna-
tional differences. It turned out that economists and policymakers were
deceived into believing that the problem of underdevelopment could be
solved and so easily. Albert O. Hirschman very succinctly sums up the
mood of disenchantment in the following.

"A drastic transvaluation of values is in process in the study of economic
and political development. It has been forced upon us by a series of
disasters that have occurred in countries in which development seemed
to be vigorously underway. . . . As a result, one reads with increasing
frequency pronouncements about the bankruptcy of the 'old' development
economics, with its accent on growth rates, industrialization, and inter-
national assistance, and about the need for a wholly new doctrine that
would emphasize income distribution, employment, and self-reliance" (p.
83).

The colloquium collection includes papers by Dudley Seers, Paul Streeten, and Fernando Henrique Cardoso, who review the literature and the past and recent perceptions of the development process. With the exception of Cardoso, who demonstrates the existence of a truly Latin American economic thought which generated ideas and motivated developmental action, most of the material in this section, while useful, is not new. Originally, the organizers of the collection wanted to confront the standpoints of the developed and the developing countries and to seek the contributions of different disciplines; only the first objective was met because four out of the 12 participants were Third World experts.

For Nurul Islam, the "revolt of the periphery" implies neither a uniform ideology nor a similar class structure; actually, the diversity of Third World societies is increasing under the impact of growing differentiation in the economic field; the common denominator remains the demand for a new international economic order. The papers dealing with the developed world's response to these aspirations reflect the mighty resistance to substantive change. Here not much has changed; one still finds such arguments as: since Singapore did it, why cannot Egypt and India!

Some of the papers are full of the kind of sweeping and vacuous generalizations that appear in the writing of UN reports; I am referring, for instance, to Goran Ohlin's statements on the military; "They have the guns, they alone incarnate national unity" (p. 129). How real people face the situation is rarely taken into account. At one point Fawzi Mansour raises a very important problem, that of regimes changing their economic policies; for example, "Egypt moving from bureaucratic capitalism to liberal capitalism while India was moving under Indira Gandhi in the opposite direction" (p. 221). While the author recognizes the importance of explaining these shifts, he decides that "it cannot at present receive the attention it deserves" (p. 221); too much space had been, of course, devoted to the "successive modes of surplus extraction . . . in human history" (p. 200).

There is no evidence of any confrontation and debate between the various points of view; as in the North-South dialogue, the papers offer a juxtaposition of ideas, not the collective effort to elaborate a common strategy. As a matter of fact, the title of the book does not correspond to its content; the colloquium must have been a useful opportunity to put things into perspective, to look back, and perhaps to learn from past mistakes, but it has no development strategy to offer. Then again, as Myrdal remarked a long time ago, rarely has the development of economics blazed the way to new perspectives on its own; more often than not, the cue to reorientation has come from the sphere of politics.

The Politics of Workers' Participation: The Peruvian Approach in Comparative Perspective. By Evelyne Huber Stephens. New York: Academic Press, 1980. Pp. xvi + 290. $29.50.

Linda Fuller
University of California at Berkeley

The most common approach to studying workers' participation has been to quarantine the workplace before examining it. An important strength of Evelyne Stephens's book, *The Politics of Workers' Participation,* is her rejection of this method and her insistence on viewing workers' participation schemes in the context of broader social, economic, and political environments. This approach allows Stephens to extend her discussion of the causes and effects of workers' participation beyond the unavoidably narrow limits of artificial isolation from surroundings. She does not mention such standard explanations of the origins and structures of workers' participation schemes as type of technology, factory size, or management style. Nor does she look at the effects of participation in terms of workers' behavior or a firm's profit or investment levels. What she does instead, by leaving the workplace in context, is recognize that workers' participation is primarily a political issue. She claims that, because of its effect on the distribution of power and material resources among different social classes, workers' participation will always be the object of political struggles at both entrepreneurial and societal levels.

Stephens investigates the politics of workers' participation in six countries which she says approximate three types of "politico-economic systems." France, West Germany, Sweden, and Chile represent the "liberal-pluralist" type of system, Yugoslavia the "bureaucratic-centralist" type, and Peru the "authoritarian-corporatist" type. The fourth type of system, "democratic-socialist," is conceptually important, but none of Stephens's cases falls into this category. Workers' participation schemes in the studied three types of countries are compared on four "developmental" dimensions: origins, purpose and design, development and effects, and linkages to policies and organizations at the national level.

A number of interesting points emerge from this analysis. The Peruvian case, in particular, demonstrates quite clearly how workers' participation schemes, even when their explicit intent is to promote identification with the firm and acquiescence to a given social system, can in fact spur the working class to mobilize around demands for both expansion of participation in the workplace and for fundamental changes in the social order. Stephens recounts in detail how this happened in Peru. In 1970, a law was passed requiring private firms to distribute 10% of their profits to

individual workers and 15% to a collective of workers, the *Comunidad Industrial* (CI). The CI was to reinvest this 15% in the firm and be granted seats on the board of directors in proportion to the amount of stock owned. Even though this opportunity for participation was very limited, less than six months after passage of the law the individual CIs around the country had begun to organize into a national federation which called for the socialization of the means of production, worker participation in state decision-making bodies, and the extension of worker participation at all levels. Stephens notes other effects of the scheme: the number of recognized unions increased dramatically following the legislation, and strike activity swelled during the same period.

Even though she shows minimal changes in relationships at the *workplace level* having a demonstrable effect on the relative strength of the working class vis-à-vis the bourgeoisie at the *national level* in Peru, Stephens's discussion highlights a critical point about workers' participation schemes; if there is no strong and unified working-class political organization at the national level, they are extremely vulnerable. In Peru, there was no such organization, and the demise of workers' participation was quite rapid. In West Germany and France, for similar reasons, the development of workers' participation has been thwarted since its inception. Stephens attributes the relatively successful history of workers' participation in Sweden largely to the strength of the Social Democrats there. In other words, the comparative data indicate that, where the working class does not control the state, the success of even limited workers' participation schemes depends on the full support and protection of a powerful working-class political organization operating in the national political arena.

But even when the working class wields considerable national political power, there is no guarantee of workplace democracy. The full development of workers' participation requires an independent working-class organization that mobilizes factory workers to take full advantage of existing possibilities for participation, and, more important, seize every opportunity to press for the expansion of the participation system. The existence of such unions has been partly responsible for the steady expansion of workplace democracy in Sweden, as compared with its stagnation in France. And, Stephens argues, even in Yugoslavia, where *formally* workers' participation is more extensive than anywhere else in the world, the lack of an independent workers' organization at the base has stopped workers from taking full advantage of structurally given participation possibilities. Without the presence of an active mobilizing agency at the factory level, workers' participation schemes will never amount to as much in practice as they do on paper.

Though her perspective allows her to make some interesting obser-

vations, Stephens's work has some major shortcomings that can be traced to the "politico-economic" classification scheme she uses. This scheme, which sets the framework for her entire analysis, is based on two types of distinctions: economic (socialism vs. capitalism), and political (polyarchic vs. nonpolyarchic). The most obvious objection is that the difference between capitalism and socialism cannot be reduced to a set of economic items like "public" versus "private" ownership of the means of production. But equally problematic, the distinction between "polyarchic" and "nonpolyarchic" political systems is based on an extremely limited notion of democracy at the societal level. This might have been obvious had Stephens compared the notion of *societal level* democracy, incorporated in the "polyarchic/nonpolyarchic" distinction, with her own discussion of democracy at the *workplace level*. Using Western "democracies" as a reference point, the "polyarchic/nonpolyarchic" distinction stresses the presence or absence of certain *oppositional* rights, for example, "freedom of expression," "right to vote," "free and fair elections," free press, etc. Workplace democracy, on the other hand, focuses on certain *participatory* rights, for example, percentage of worker representation on decision-making bodies, direction of worker influence (veto power vs. power to initiate discussion of issues), types of decisions workers participate in, etc.

The limited meanings of democracy, socialism, and capitalism imposed by this classification scheme weaken and confound Stephens's analysis of the politics of workers' participation. She is forced into the untenable position of viewing democratic socialism as merely a state-owned economy grafted onto a bourgeois democratic polity. Since she argues that full development of workers' participation depends on a successful transition to democratic socialism, she can then conclude that what is needed in "liberal-pluralist" countries is a change from a capitalist to a socialist economy. The polity is already "polyarchic" and must remain so. In other words, working-class strategy for democratizing their workplaces consists of voting capitalism out of power in "free and fair elections." Even Stephens admits in her conclusions that she is not very optimistic that this will ever happen.

Equally detrimental to the analysis is the lack of a clear understanding of the differences among the political, economic, and social relationships of developing and developed countries. The fourfold scheme Stephens uses to guide her comparative investigation does not even distinguish between the two. And while the terms "developing" and "developed" are inserted into her discussion at many points, they do not appear to refer to much more than a difference in the amount of "resources" available in one country and not another. But they refer to much more. They refer to the centuries-old development of capitalism on a global scale that

resulted in the division of the world into developed and developing areas, each with distinctive economic, political, and social relationships.

This lack of a full understanding of underdevelopment is apparent throughout Stephens's discussion of workers' participation in Peru. Students of underdevelopment have long warned that the independence and autonomy of the state in developing countries are an illusion in the sense that they reflect the weakness of internal social classes rather than the strength of the state itself. Nevertheless, Stephens evidently takes appearance for fact. As a result, her analysis of the political dynamics surrounding the origins, development, and demise of workers' participation overemphasizes the importance of the actions, ideology, and composition of the Peruvian political "elite." Thus we learn that internal divisions within the government were largely responsible for the compromise character of the workers' participation design, and that the eventual scrapping of the whole idea was a result of the junta's lack of a "clear vision," "ideological consensus," and "coherent strategy" which rendered it "susceptible to outside influence of capital-owning groups" (p. 235). What this interpretation lacks and what a firm grounding in the development literature could have provided are a realistic sense of the obstacles placed by underdevelopment not only to any effort, like that of the Peruvian junta, to reform capitalism, but toward the effecting of a peaceful transition to democratic socialism where workers' participation could someday flourish.

"We Come to Object": The Peasants of Morelos and the National State. By Arturo Warman. Translated by Stephen K. Ault. Baltimore: Johns Hopkins University Press, 1980. Pp. 319. $20.00.

Kennieth Pittman
Manhattanville College

"We Come to Object" is a product of the Seminar on Peasant Societies at Mexico's Center for Advanced Research of the National Institute of Anthropology and History. In the only work of the group that has been translated into English, Arturo Warman, the seminar's director, attempts to clarify the sources of structural change in Mexico and "who the strongest, most important protagonists in this process are" (p. 3). And, like Barrington Moore and others before him, Warman points to two principal actors: the peasants and the State. It is these two who "are the bearers of the most critical contradiction, central in the process of a capitalist industrialization that is dependent and tardy, if not, indeed, posthumous and unfinished" (p. 4). In chapter 1, Warman selects the eastern region

of the state of Morelos to serve as a microcosm of the national experience. The following chapters chart the evolution of Mexico's particular brand of dependent development.

When sugarcane became the mainstay of the local economy in the late 16th century, the mechanisms of surplus expropriation which were to remain fundamentally unchanged for the next 400 years were firmly in place. A monetarized colonial economy sat astride a preexisting subsistence economy. Capital investment in other sectors was generated from the surplus expropriated from the peasantry. The hacienda (plantation) clearly had its roots in the colonial era, but it was, above all, the child of Mexico's 19th-century liberal capitalism. Under the Díaz dictatorship, eastern Morelos became one vast hacienda. By separating labor from the land, the estate owner could both force the peasant to reproduce himself in the subsistence economy and require that he pay the capitalist sector over him for the right to subsist.

The Zapatista rebellion of 1911 was witness both to the erosion of the balance between capitalist and subsistence sectors and to the inability of the Porfirian State to mediate in the exchange. Acting in a defensive maneuver to forestall the collapse of the equilibrium between peasant and capitalist sectors, the Zapatistas broke down the old structures of domination. Their efforts to reconstruct a State in their own image made their rebellion into a revolution. The Zapatistas never attempted to capture the State but instead to transform it in an effort to go on living as peasants.

It is in this revolutionary conflict between the peasants' vision of a decentralized, protective State and the modernizers' vision of a centralized, extractive State that Warman locates the critical characteristic and contradiction of Mexico's dependent development in the 20th century. The centralized State was ultimately victorious, but it emerged in the postrevolutionary period as an independent entity unattached to any class. Adopting the project of national industrialization as its own and distanced from the contradictions between classes, the State assumed a directive role. Agrarian reform put the Zapatista revolution at the service of the State and guaranteed the perpetuation of a peasant sector capable of producing a surplus to be transferred to industry.

Since 1930 Mexican development has been grounded in an intensified version of those same mechanisms of expropriation that stood at the core of the colonial economy. But the expropriation of the peasant surplus is now sanctioned by a State which has assumed a participatory role in the process of industrialization. The State presides over the intermediation between industry and agriculture and extracts the peasant surplus by allowing the cost of subsistence, the price of survival, to rise. The "true miracle of Mexican development," says Warman, is located in the peas-

253

antry, and it is located there not despite its poverty, "but more precisely, because of it" (p. 253).

But, if the maintenance of a peasant economy at the center of a capitalist development is the essential characteristic of the Mexican experience in the 20th century, it is, for that reason, its critical contradiction. The State has become the responsible agent, and, from the perspective of the peasants, has provided survival at the cost of stretching intensification over fewer and fewer resources. According to the developmental logic adopted by the State, the capitalization of the countryside and its peasants is ultimately necessary, but therein lies the paradox. Faced with destruction, the peasants could be counted on to, as they have over the centuries, "come to object." The State would be then forced to confront the destruction of the very sector on which its dependent industrialization is founded.

Warman's thesis is not new; he makes no claim as to its novelty. His presentation of the evidence from eastern Morelos, however, forms a compelling argument and is a genuine contribution to the literature of dependent development. One can question the extent to which the State is actually as autonomous as Warman seems to claim, but his analysis leaves little doubt about the nature of the contradiction that State confronts.

Revolt against the Dead: The Modernization of a Mayan Community in the Highlands of Guatemala. By Douglas E. Brintnall. New York: Gordon & Breach, 1979. Pp. xxiii + 199. $21.00.

Kay Barbara Warren
Mount Holyoke College

Studies of ethnic relations have been among the most significant products of cultural anthropology in the Mesoamerican region of southern Mexico and the Guatemalan highlands. Douglas E. Brintnall's *Revolt against the Dead* makes an important contribution to this literature from a materialist perspective. This study of Aguacatan, a rural *municipio* (county) located in the northwest highlands of Guatemala, is particularly interesting because of the ethnic heterogeneity of the region.

In contrast with the findings of many community studies in Guatemala, Brintnall finds that the peasant Indian/Ladino elite cleavage is not the only relevant ethnic division. Rather, in Aguacatan, subdivisions of the *Indian* population (into Easterners, Westerners, Quiches, and Mam) have played an important role in the recent collapse of traditional Indian politics and religion and have posed new challenges to the Ladino domination that had been maintained previously by their superior commercial position

and monopolization of the brokerage positions between rural communities and national society.

Brintnall's analysis is most incisive when he focuses on the economic transformation of Aguacatan, first through changing forced labor policies which compelled Indians to work on distant coffee plantations, and, more recently, through the introduction of a lucrative cash crop, garlic, which has reversed the impoverishment of those Indians who have access to irrigated croplands. His subtle socioeconomic analysis explores a terrain—the changing internal stratification of ethnic groups—that many students of interethnic relations have just discovered.

The author finds that the major Indian ethnic groups live in different ecological zones in the county and practice distinct forms of economic production. The Quiches specialize in higher-altitude crops, livestock, and lumbering, while the Eastern and Western Aguacatecs, who inhabit a lower zone, cultivate corn, black beans, and irrigated cash crops. This variation in modes of production becomes important in the explanation of the differential participation of Indians in a rebellion against traditional religion in the 1950s and 1960s. Those who lived in areas of irrigation and cash cropping were much more likely to sever their ties to traditional Indian religion and to convert to the new religious groups proselytizing in the area. Patterns of conversion show a socioeconomic basis, with the Catholics tending to receive the wealthier converts and the Protestants those poorer Indians who still depended on seasonal migratory labor for their livelihood. In Aguacatan these new religious groups are part of a tidal wave of new institutions—such as agricultural cooperatives, political parties, and peasant leagues—that have destroyed the authority of traditional Indian leadership, tied Indians more firmly to national life, and undermined the control of local Ladino elites.

What Brintnall does so well is to show that the political enactment of change in the daily lives of the people of Aguacatan does not involve the definitive substitution of new cleavages for old. Nor does the proliferation of new local affiliates of national organizations mean that older, localized agendas are totally swept aside. We are shown the new complexities of local factionalism that leads to new forms of interethnic Indian cooperation and joint action, while primordial ethnic divisions still reassert themselves from time to time. Brintnall argues that growing class divisions among the Indians (resulting from differential access to cash-crop production) are implicit in some of the new institutions, but that class antagonism is muted by a dominating concern with challenging local Ladino domination.

Brintnall's book is clearly written and painstakingly crafted. It would make a good case study for advanced undergraduate and graduate courses because of its sustained analytical argument and ethnographic compre-

hensiveness. There are, however, several puzzling shortcomings in this work. Brintnall discusses social organizations and patterns of behavior, but he gives us only brief glimpses into the cultural shapes of Indian world views, value systems, and social ideologies. We are never given a complete sense of *their* frames of reference for comprehending the changes they are living with. Brintnall's tendency to reduce Indian religious belief to a concern with "the wrathful power of the dead" contributes to this omission. Without fuller data on the myths, rituals, and religious beliefs that give cultural form to Indian *costumbre,* we are unable to ask such questions as: Did "traditional" Indian religions maintain significant differences in world views among Indian ethnic groups on the eve of renewed state expansion? Did religion serve as an idiom for the expression of Indian understandings of their exploitation by Ladinos? Did changes in religious affiliation entail transformations in the cultural categories through which Indians created their understandings of the politics of their social world?

Another weakness is in the time frame of the book. The author has selected the Ubico regime of 1933–44 as a preeminent representative of the past. No doubt this was a watershed period of state expansion, but one is left wondering whether Indian and Ladino oral histories incorporate a greater time depth that would allow a fuller analysis of the changing perceptions of national society.

Revolt against the Dead is a very successful materialist analysis of Indian ethnicity and social change in Mesoamerica. We are still waiting, however, for an ethnographer to do equivalent analytic justice to the socioeconomic and symbolic dimensions of Indian identities in the context of rapid social change. The Brintnall book does move us a step closer by helping to reconceptualize our approaches to ethnic cleavages, economic stratification, and social change in Guatemala.

The Politics of Domesticity: Women, Evangelism and Temperance in Nineteenth Century America. By Barbara Leslie Epstein. Middletown, Conn.: Wesleyan University Press, 1981. Pp. xli + 188. $17.95.

Access to Power: Cross-national Studies of Women and Elites. Edited by Cynthia Fuchs Epstein and Rose Laub Coser. London: George Allen & Unwin, 1981. Pp. x + 259. $27.50.

Elise Boulding
Dartmouth College

Barbara Epstein's historical treatment of 19th-century evangelism as the politics of domesticity in *The Politics of Domesticity* and Cynthia Epstein

and Rose Coser's comparative sociopolitical studies of 20th-century women's access to power in *Access to Power* provide, each from its own disciplinary perspective, an interesting documentary of the primary concerns of many contemporary feminist scholars. Through the documentation of the structural placement of women and men in society and of how existing male dominance structures continue to track women away from significant decision-making roles whether in the domestic, religious, economic, or political sphere, women's powerlessness is presented to us as primarily a product of the historical forces at work in the modern era.

Epstein's treatment of women's participation in both the 19th-century religious awakenings and the temperance movement which grew from that same evangelistic soil is squarely set in the context of an urban male dominance–female dependency phenomenon that replaced a more egalitarian male-female relationship rooted in a more rural economy. Although it was institutionally under male control, evangelism alternately served as an emotional solvent for female rebelliousness and as a golden opportunity for controlling husbands by praying for them publicly. Epstein arranges her historical materials very ingeniously to show a progression from (1) the 1740s religious awakening in a still-rural society that was experienced and reported by men and women in rather similar ways, to (2) the turn-of-the-century religious awakening that was characterized by growing, but masked, sexual antagonism, on through (3) the 1870s "Ohio Crusade" when women's religious fervor generated startling new courage for women's public activism—albeit conducted with a traditional, self-effacing personal demeanor—and finally to (4) the emergence of the proto-feminist Women's Christian Temperance Union (WCTU) which somewhat gingerly supported equal rights for women as part of its other reform goals.

Epstein points out correctly that women's culture tended to abjure the discussion of power for the discussion of morality, although women's actions were *shaped* by considerations of power. What is missing in this otherwise skillful and illuminating historical analysis is discussion of the more encompassing social goals which a significant number of 19th-century women of the "middling classes"—Epstein's target population—envisioned and worked for. Much of her documentary analysis is of more narrowly religious writing by men about women, and of men's selections of women's writings, and so she misses the power of the great 19th-century women religious activists. Women missionaries were the first to articulate the theme, "women's work for women." They were on every continent, representing every denomination from "traditional" Catholicism to ardent, evangelistic, out-of-church movements, from 1800 to 1900. There is no mention of the female utopian religious leaders who played a major

public role in the second quarter of the 19th century. Ellen H. White of the Seventh Day Adventists, Mary Baker Eddy, founder of Christian Science, Ann White, the Shaker elderess, feminist, and pacifist, and other less well-known women, all were concerned, in some degree, with re-shaping the existing social order and elucidating the relationship between morality and power.

Epstein misses the fact that the pentecostal power of the Ohio Crusade of the 1870s fueled the international women's peace movement in the 1880s and continued to do so when former crusaders like Mary Brayton Woodbridge turned to the study of statecraft and arbitration processes in order to deal with the larger problem of war. The treatment of the WCTU itself also diminishes the stature of the women who made it the first international organization to link international problem solving through mediation with domestic urban poverty and the economic and political status of women, all within the broader framework of a new human morality. Temperance was only one aspect of that program, and it was not pursued simplistically. What is extraordinary is not that so many women accepted male-controlled religious institutions or used religious movements to compensate for their powerlessness, but that some women had visions of an entirely different social order, with different roles and relationships among women and men, and, from their powerlessness, developed some skills, however limited, in working for social transformation.

Cynthia Epstein and Rose Coser, pioneer feminist investigators of our own times, bring a rich experience of theory building and empirical research to their task of organizing comparative studies of the power, and powerlessness, of women in elite decision-making careers. Drawing largely on fellow sociologists, they have nevertheless brought together an interesting diversity of social science expertise, including law and comparative politics and management from six Western European countries, two Eastern European countries, and the United States. The studies focus on the consequences of women's differential location in the social structure and confirm previous findings that women cluster in certain occupations where, as Coser says, "like the sediment of a good wine, they have sunk to the bottom" (p. 16). Few women in any country rise to positions of decision-making power. The general findings of the studies are well summarized in an introductory chapter by Epstein.

International symposia are difficult to organize and materials are rarely as neatly comparative as one would like. National "stories" on a common theme are interesting up to a point, but one must go beyond this. The present book suffers from a dearth of directly comparative material, but what there is, is very interesting. C. B. Silver's comparison of the status

of the public and the private sectors in France and the United States, and her findings that women in each country rise higher in the sector with the lower status, are a model of comparative research. The observations on the unique impact of German culture on women's status (D. S. Sanzone, H. Pross) suggest an important comparative theme that I wish had been developed more consistently throughout the book. The accounts of the impact on women of the interaction between culture and ideology in Poland (M. Sokolowska) and Yugoslavia (B. Denitch) (ideologies can "melt" social structures) are of great interest, and one theme continually recurs throughout the book: intervention strategies and the political will to intervene do make a difference in giving women access to power. The concept of social capital and the analysis of the forms available to women and to men and the different ways of accumulating it are valuable contributions of the book. The chapters on Austria (H. Nowotny), France and Britain (R. Whitly), and the United Nations (B. Thom) use this concept very productively. Also interesting is the concept of new achievement tracks for women in which professional competence may override sex handicaps, specifically through the acquisition of law and MBA degrees—possibly also through computer competence. This theme appears in several of the nation studies in part 2. The studies also indicate, however, that the old roads of voluntary sector experience and widowhood are still main roads to power for women.

Interesting as all these findings are, the basic concept of "access to power" remains a rather thin one. As several of the studies point out, women, when placed in similar positions in the role structure, think like men. Power for women in male power structures then becomes an end in itself. The question of what a preferable future social order that gives new roles to both men and women might look like is by definition unaskable in the "access to power" framework. The voluntary sector is usually where women's social change energies operate. Epstein makes some interesting observations on the voluntary sector in her chapter on United States women in politics, but the larger issue of women's power to reconstruct the social order is deferred. To do justice to that subject would, of course, require a different kind of book.

The Politics of Domesticity and *Access to Power* both reveal the analytic power of feminist research strategies. They are both important contributions to the study of social power. The two books also reveal the limits of current research formulations about the role women play in the social order. I hope the next round of feminist studies of power will be on the theme: women's power—for what?

The Female World. By Jessie Bernard. New York: Free Press, 1981. Pp. x+614. $19.95.

Ruth Shonle Cavan
Northern Illinois University

Jessie Bernard opens *The Female World* with this statement: "The idea is to deal with the female world in and of itself, as an entity in its own right, not as a by-product of the male world" (p. 3). The book is for the general reader, not to entertain but to enlighten about a specific feminist point of view.

The book is encyclopedic in coverage. Bernard establishes that there is a female world among subhuman primates as well as in numerous human societies at different periods of time. She then discusses past stations and spheres of women, the incomplete demise of a distinct women's sphere, the age and social class structure of the women's world, networks and associations, culture and ethos of the female world, economic and political status, and finally the relationship between the female and the male worlds.

Although Bernard's stated purpose is to treat the female world as a distinct entity, she refers constantly and always negatively to the effect of the male world on the female world. She sets up the extremes of culture and ethos of the two worlds. The female world is gemeinschaft in nature with either a love or a duty ethos or both, and is expressed in filial duty, concern for children, and public humanitarianism. In contrast, the male ethos is based on a gesellschaft relationship and includes an aggressive, even violent, stance and is dominated by self-interest. In reference to the relationship between the female and male worlds, Bernard states (p. 11) "... although individual men may love individual women with great depth and devotion, the male world as a whole does not. Terms like 'hatred' or 'hostility' may be too strong to describe this response, so the term 'misogyny' has been invented." The word "misogyny," which appears 11 times in the index, is defined in dictionaries as "hatred of women." Bernard does not discuss the vast area of behavior that lies between the extremes of female and male ethoses as she describes them.

One theme emerges constantly throughout the book. Men have dominated women, oppressed them, feared their competition, hated them. Until well into the present century the male ethos was dominant and few women made forays into the male world. However, with industrialization, better means of contraception, education, and labor-saving devices, women gained freedom and, along with it, dissatisfaction. Options for women began to appear, marriage or nonmarriage, motherhood or childlessness, dependence or equality. The struggle for equality was on.

An undertone of blame of men runs through the discussion. Bernard does not seem to see that men, as well as women, are caught in the same cultural bind from the past when roles were better coordinated, that culture changes slowly, and that many men have aided women in their struggle.

Bernard has not carried out any original research to test her assumptions. The book carries a reference list of some 800 items on which she draws to formulate and support her contentions. Selecting sources to support an assumption while neglecting negative data is always dangerous. Bernard's reputation as a sociologist is sound, but she may be swayed by her personal feminist views.

In summing up, Bernard does not expect the male ethos to continue its domination; nor does she anticipate that the female ethos will take over. Rather, she sees a softening of the male ethos through increased participation of women in economic and political activities. Instead of continued opposition between the female and male worlds, she anticipates integration.

The book is well worth reading for its historical tracing of women's struggle for recognition by the male world and for its exposition of the current feminist viewpoint on the relationship between female and male worlds.

Women in Law. By Cynthia Fuchs Epstein. New York: Basic Books, 1981. Pp. x + 438. $18.50.

Robert L. Nelson
American Bar Foundation and *Northwestern University*

Cynthia Fuchs Epstein began studying women lawyers when she was a graduate student at Columbia, interviewing a small random sample of New York attorneys in 1965–66. Her object was to examine those women who made a "deviant" career choice and successfully penetrated a male-dominated profession. While Epstein pursued other research on women in the professions, the role of women in the legal profession was undergoing a profound change. Whereas in 1965 only 4% of all lawyers were female, by 1980 12% were female; in 1965, 4.3% of law students were female; by 1980, one-third were female. By the end of the 1970s, women lawyers had begun to enter the previously all-male bastions of the profession—the large law firms, the committees of professional associations, and the judiciary. (The first woman Supreme Court justice, who was appointed in 1981, was offered nothing but secretarial positions in major West Coast law firms in 1952 after her graduation from Stanford Law

261

School where she was an editor of its law review.) By the time Epstein finally published her dissertation, her focus had shifted from explaining deviant career choices to explaining the causes, consequences, and continuing frustrations of the women who are now entering the legal profession in unprecedented numbers. Her analysis of the halting and variable success of this transformation and the historical context of sex discrimination in legal institutions in *Women in Law* makes an important contribution toward a systematic sociological treatment of the subject.

Epstein's basic argument, introduced in the first chapter, is that the legal profession offered new opportunities for women during the 1970s as a result of a dramatic rise in the number of lawyers (43% from 1970 to 1979), the increased demand for associates in large law firms, the impact of civil rights legislation on firms and law schools, the growth of public interest alternatives to corporate law practice, and the increased activism by women's groups inside professional associations and law schools. This ferment spurred a significant increase in the proportion of women applying for and gaining admission to law schools. So positioned, women began to pursue a full range of legal careers despite continuing tensions produced by sex stereotypes, the competing demands of family relationships, and a hypothesized " fear of success." Epstein argues that programs to increase opportunity for women (such as special recruitment by law schools) tapped a reserve of female interest in the legal profession, altered the career attitudes of women entering law, and began to alter the perceptions of male colleagues and spouses about the appropriateness of women's pursuing legal careers. According to Epstein, such social psychological factors as women's self-images and the roles women play in the family are far less limiting than some scholars of sex roles have suggested. She asserts constantly that attitudes and relationships are perceptibly different from the dark days when she began her research, and that further enhancement of opportunities is likely to lead to further participation by women.

In advancing this proposition, Epstein examines a broad range of professional contexts, beginning with the gatekeepers of the profession, the law schools, and then devoting a series of chapters to various practice settings, law school teaching, the judiciary, and professional associations. She concludes with five chapters dealing with the dynamics of women's relationships with colleagues, spouses, and children. Throughout the book the analysis shifts between data on the changing structural conditions of women in law and qualitative data on the social psychological aspects of their involvement in a male-dominated profession. There are recurrent discussions of role models, of reactions and feared overreactions to men, and of the value conflicts between seeking the conventional rewards of a professional career (income, status) and remaining faithful to the social values of the women's movement. In addressing a range of questions in

a range of legal contexts, Epstein produces a book of broad interest and appeal. The separate treatment of various segments of law practice is a realistic presentation of the highly stratified nature of the profession, and many of the detailed findings (such as the data on the social backgrounds of women lawyers) will interest students of the legal profession. The chief limitation of Epstein's approach is the lack of systematic data collection during the 1970s. She does not offer any set of in-depth interviews with a random sample of lawyers comparable with her original interviews. Instead she takes data from a "snowball" sample of lawyers practicing in different locales and different settings who were interviewed at different times. Many of these were met through contacts in the women's movement. In going beyond the aggregate data on structural change, we must rely on Epstein's personal observations of change and quotations from interviews that are not, on their face, probative. The analysis of changing values and relationships is plausible, interesting, and clearly enriched by the lengthy exposure of the author to women lawyers, but it suffers unnecessarily from casual sampling and presentation.

Some chapters are so brief that they serve primarily to touch base with aspects of the legal profession; others provide greater depth. The discussion of the rise and fall of feminist law firms (chap. 9) reveals aspects of the "market" for feminist legal services, and the pressures such organizations face because of cash-flow difficulties, competition from family needs, and ideological conflicts. The analysis of the entry of women into large law firms (chap. 11) is interesting both for its study of rapid organizational change and for the questions it raises about the prospects for women's gaining partnership, client responsibility, and leadership roles. The analysis of the ambivalence experienced by women lawyers (chap. 15) is an excellent general discussion of relationships between the sexes in a professional, not just legal, context. The chapters examining the link between women's roles at work and in the family (chaps. 17, 18, and 19) testify to the liabilities women have faced traditionally as a result of male dominance in the family. In 1970, women lawyers averaged seven fewer hours of work a week than men, were three times as likely to work part-time, were four times more likely to forgo marriage, and if they married and reared children, were likely to shoulder primary responsibility for domestic chores and child care. As Epstein suggests, how these patterns change and how women adjust to changing roles in work and the family are critical to the long-run participation of women in law.

Women in Law is an important and well-written account that will be of interest to scholars of sex roles and the legal profession alike. I hope that the important questions it raises receive further systematic analysis in the future.

Press and Public: Who Reads What, When, Where, and Why in American Newspapers. By Leo Bogart. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1981. Pp. xviii + 285. $19.95.

David Weaver
Indiana University

Press and Public is a thorough and thoughtful report on newspaper readers and changes in their tastes and habits. It is based on (mostly) national studies of readers conducted under the auspices of the Newspaper Advertising Bureau during the past two decades. Most of the approximately 27 studies that provide the foundation for this book employed samples of 1,000 or more, and many used samples of several thousand in size.

Leo Bogart, now executive vice president and general manager of the Newspaper Advertising Bureau, has been an advocate of the newspaper industry since he joined the bureau in 1960 and is, in his own words, "a hired hand" (p. xiii) of the newspaper establishment. However, the 137 tables and three charts in this book contain a great deal of timely and valuable national data on newspaper circulation trends, population mobility, advertising expenditures, newspaper content, newsprint prices, newspaper readership (by days, time of day, age, location of residence, education, income, sex, race, size of community, readers' use of other media, political participation, marital status, personal mobility, and different types of newspaper content), television viewing, topics of discussion, attitudes toward newspapers, television news content, editors' perceptions of editorial quality, kinds of news most remembered, and interest in various kinds of newspaper content.

Bogart's analyses of these data are also valuable, especially in the chapters dealing with the various challenges to newspapers, the nature of news, and newspaper editors' views as compared with readers' views. But there is little reference to theories of newspaper readership and use, such as those proposed in the various media "uses and gratifications" studies so popular in the past decade. In fact, the book as a whole is surprisingly devoid of discussion of theory, opting to remain at a mostly descriptive level, largely, one suspects, because of the descriptive nature of the data undergirding it. Thus the major strength of the book is the provision of large amounts of specific national data on newspaper audiences and content, rather than a more general theoretical treatment of the subjects. Because of its encyclopedic nature, the book serves more as a valuable reference tool than as one that is compelling reading. In this respect, Bogart's book is very like the John Johnstone, Edward Slawski, and William Bowman book, *The News People* (Urbana: University of

Illinois Press, 1976), which provides extensive national baseline data on journalists and their work.

Bogart is careful to point out that the book is *not* a comprehensive review of the newspaper readership research literature and that the generalizations from national studies cannot be transferred automatically to individual newspapers in local settings. Although I sometimes wish that the author had made a greater effort to relate the many findings of the bureau's national studies to other more local readership research regularly reported in *Journalism Quarterly,* the *Newspaper Research Journal,* and the American Newspaper Publishers Association *News Research Reports,* there is no denying that this book is a gold mine of national newspaper readership and content data. To have many of the various reports, papers, and presentations of the Newspaper Advertising Bureau collected in one well-organized volume is extremely valuable, provided one remembers that the data and the conclusions are those of a promoter of American newspapers.

This is not to imply that these data, or interpretations of them, are misleading, but only to make the point that this is not a critical review of U.S. newspaper content or readership. It is also not an openly biased account, although it often does tend to use the data to see a silver lining for newspapers. In the final chapter, for example, Bogart concludes that "the American press has never been more prosperous or more essential to the political health of the nation" (p. 254). But he also cautions that "it is difficult for a monopoly daily to avoid complacency and establishmentarianism" (p. 264).

The book concludes with a fairly skimpy section on newspapers and new technology, in which Bogart argues that teletext and viewdata systems are not likely to make the newspaper obsolete (a view I share) but are likely to have "remarkable effects on newspaper content" (p. 270). But Bogart's true colors show in the last paragraph of the book, when he asserts that "no technological change can ever challenge the newspaper's command of big ideas, its traditions of deep inquiry, sweeping synthesis, and inspired advocacy" (p. 272).

In spite of this pronewspaper point of view, *Press and Public* is a remarkably rich source of data on U.S. newspapers and their readers which is carefully interpreted by a veteran Chicago-trained sociologist who makes a genuine effort to be fair and evenhanded. No one seriously interested in systematic study of U.S. newspapers and their audiences can afford to ignore this benchmark work.

Robert W. Kubey
University of Chicago

The TV Ritual, by Gregor Goethals, is a valuable attempt to understand television through its presentation of ritual and iconic images thought, in this analysis, to help maintain social cohesion. The author's central claim is that television content presents a mythical "symbolic universe" that shelters its viewers from anomie and provides a means of location in the world. Goethals documents TV's potential to "substitute for sacraments" in illustrations of how the medium allows people throughout the world to be present at major ritual events. Goethals cites Pope John Paul II's installation, John Kennedy's funeral, and Ronald Reagan's inauguration as examples. The height of secular ritual on American TV, for Goethals, is the Super Bowl where "Super Sunday" becomes a "sacramental occasion."

Icons are defined as symbolic forms used for the objectification of beliefs and self-transcendence. An icon brings persons into relation with the sacred through its visual narration of sacred stories. In American culture, according to Goethals, there are three major iconic stories: (1) community life (family, neighborhood, city, nation); (2) nature; and (3) technology. She claims, as has Horace Newcomb in *TV: The Most Popular Art* (Garden City, N.Y.: Anchor Books, 1974), that these stories were not newly created for television but hark back to older American cultural symbols.

In discussing family icons, Goethals analyzes "All in the Family," "Eight Is Enough," "One Day at a Time," "Dallas," and "The Waltons." Her examination of nature concentrates on "Little House on the Prairie," "Grizzly Adams," and facets of the Western genre. Again, she concurs with Newcomb that a central theme in Westerns is the creation of order in an untamed land. The "machine" programs discussed are "Chips," "Charlie's Angels," "The Dukes of Hazzard," and "Star Trek." In the chapter on TV as "iconoclast," she offers examples of early TV documentaries such as "Harvest of Shame" as well as "Sixty Minutes," "Saturday Night Live," and the Senate Watergate hearings.

Goethals presents the overall theoretical premise early in the book and, while the descriptions of familiar television fare are interesting and occasionally illuminating, they are not well integrated with the major theses. Greater articulation between theory and example would have produced a more satisfactory result than this work, which lacks consistent rigor from beginning to end. Furthermore, Goethals failed to incorporate, much

less cite, the seminal contribution of Jean Cazeneuve ("Television as a Functional Alternative to Traditional Sources of Need Satisfaction," in *The Uses of Mass Communications*, edited by J. Blumler and E. Katz. Sage Annual Reviews of Communications Research, vol. 3 [Beverly Hills, Calif.: Sage, 1974]), who has also compared the functions of television with those of myth, ritual, magic, and religion. Still, Goethals's book is a welcome and thought-provoking analysis of what, heretofore, has remained a nearly uncharted area of inquiry.

Media, Ideologies and the Revolutionary Movement. By Armand Mattelart. Atlantic Highlands, N.J.: Humanities Press, 1981. Pp. xxvi + 258. $32.50.

Daryl Hobbs
University of Missouri

Media, Ideologies and the Revolutionary Movement is a strategy book that is the work of a published scholar in the service of a cause. The book's four essays were written during the 1970–73 ideological and political struggles of the Allende government in Chile. Because the author was both advocate and participant, there is little of the analytical detachment of the scholarly work that is typically produced from a distance in time, or place, or both. The work is scholarly, but it is not detached: it reflects some of the emotion of the struggle, a commitment to "winning" it, and the uncertainties associated with the author's knowing neither the outcomes nor what would produce them. Armand Mattelart makes his allegiances clear by repeated references to "we" and "the enemy." The essays were written to analyze the role mass media play in reinforcing an existing class structure, and from that, to conceptualize and outline a mass communications strategy for the Left.

In the first essay, the mass media are seen as a driving force of imperialism (both internal and external) and technocracy which perpetuate dependence through images and fetishes designed to individualize, impersonalize, fragment, and demobilize "mass" audiences and thus reinforce the existing class structure. Instead of leading to "objective" reporting of the news, the canons of liberal journalistic practice, according to Mattelart, are being integrated into the "totality of relations of production and social relations which must be grasped if we are to understand its function in reproducing the everyday legitimacy of bourgeois relations of domination" (p. xxv). He sees such journalistic by-words as "public opinion" as means of suppressing social movements by an impersonal glossing over of dissident elements. Solidarity is undermined by the emphasis on models of competitive and individualistic behavior. Similarly, the press

is property, and thus "freedom of the press" is tantamount to the freedom of property, representing most strongly the economic and political interests of the owners of the means of production. Furthermore, the various forms and genres of the media discourage solidarity through the segregation of audiences and the creation or reinforcement of different interests among women, youth, elderly, etc. The pluralistic images conveyed by the media discourage the emergence of common interests. The modern man, as a creation of the media, ". . . becomes a kind of koala bear who, in order to roam over the whole world, must change it and reduce it to the eucalyptus tree in which he is born and brought up, and where he will die" (p. 129).

The author suggests that, in order to overcome the dominance reinforced by the media, an alternative conception of the media should be specified. Because the media are dominated by the culture, this cannot be accomplished by a simple inversion of the political signs of the media content. The vertical structure of message creation itself must be altered. Instead of fragmenting, dividing, and impersonalizing, the media must become an agent of "conscientisation," convey images of a totality of relations, and include participation of dominated persons at all levels of message creation. Such a process is not simply a transitional stage: it must be continuous.

During its three years, the Allende government did not succeed in gaining control of the mass media. Even those media that were controlled by the Left were ineffectual agents of change, either because they were following traditional models of journalistic practice and message creation, or because the audience being appealed to found the messages dull and uninteresting. The oligarchically and externally controlled media shifted tactics effectively and became an instrument of mass mobilization against the Allende government. The media "abandoned their normal role as atomizers and demobilizers of their audiences. . . . They became agitators and collective organizers, to use Lenin's prescription for the function of the press. They set about propounding a new concept of solidarity and altering the individualistic assumptions of their political clientele" (p. xix). This shift in the function of the media was successful in politicizing many previously inactive sectors of the population and was a major contributor to the economic instability preceding the 1973 military takeover.

The book is an important addition to the literature on class and dependency analysis of information, communication, and the media because of its conceptualization and its analysis of the Chilean experience. However, in my judgment, the book also adds to the literature on advocacy and applied sociology, although the author does not state this as an intention. Its value for this purpose is as a case study, formulated in the midst of a conflict, of scholarly analysis and strategy.

Crime News and the Public. By Doris A. Graber. New York: Praeger Publishers, 1980. Pp. xviii + 239. $28.95.

Dan Schiller
Temple University

Doris A. Graber vigorously pursues three main lines of inquiry in *Crime News and the Public:* content analysis of crime and justice news, in-depth panel interviews to gauge public perceptions in this area, and sustained comparison of media accounts with audience perceptions.

The content analysis is extensive and thorough. Graber has coded and analyzed several newspapers and done highly detailed crime news coding on one, the *Chicago Tribune.* Over 19,000 stories from the *Tribune* were examined. She has also performed content analysis of local newscasts by CBS and NBC affiliate stations in Chicago and studied printed abstracts of national network newscasts. The sample for the book comprised newspapers and newscasts produced over the course of one full year (January 1976 to January 1977).

Audience panels of 48 respondents each were consulted at three locations (chosen to include residents of a small town, a middle-sized city, and a large metropolitan area)—Indianapolis, Lebanon, New Hampshire, and Evanston, Illinois. Each panelist was interviewed 10 times by telephone. A fourth "core panel" of 21 Evanston respondents were paid to provide in-depth interviews and daily diaries. Questions about many topics were so interspersed that, in the panelists' eyes, no major issue focus stood out.

Content analysis revealed a heavy news media dependence on coverage of crime and justice. The *Tribune* apparently gave over 50% of its news coverage to such topics; national television news accorded them 13%. On the latter, individual crime news alone (exclusive of other justice and crime categories) was on a par with news about the presidency and Congress and was more than twice as frequent as economic reporting. As has been observed by others, murder was heavily overrepresented in news compared with the actual police statistics on it.

So much for crime news. What of the public? Generalizing about her illuminating comparison of public perception and media portrayal, Graber claims (p. 54) that "the press does not depict criminals and victims largely as nonwhite, poor, and lower class, but the panelists do." In her perspective, the data suggest the need for modification, at least, of the agenda-setting hypothesis to take into account particular audience sensitivities, demographic features, and experiences, when it is used to trace the effects of media content.

In my opinion, Graber herself has not done so—or, rather, has not

controlled sufficiently for or explicated the social patterning of audience perception. Panelists were selected from a pool of respondents chosen randomly from lists of registered voters. Such a sample, as Graber concedes (p. 7), constitutes a "somewhat select pool"—whose perceptions and attitudes, particularly in such a sensitively barometric field as crime and justice, must be *shown* to be representative. Representativeness should not be merely assumed. Even within this already skewed sample, however, other biases are apparent. Persons in clerical and unskilled occupations constitute only 12% of the total panel sample; unemployed persons only 3% of it. College-educated persons constitute 43% of the total sample, but fully 62% of the core Evanston panel. The latter is also top-heavy with professional, managerial, and technical personnel. Again, we are informed—as if it may not matter–that only two core panelists are black; we are never told about racial characteristics in the other panels. It is therefore not at all likely, as Graber claims (p. 8), and certainly may not be assumed, that differences between panelists' and less advantaged groups' perceptions "are matters of degree rather than substance."

Differences in panel perceptions which do emerge, despite the skewing of the sample—between men and women, black and white respondents, and panelists from different communities—suggest strongly that the author should have investigated the social patterning of audience perceptions more carefully before concluding that "the most important actors in determining the impact of crime and justice news are the millions of individual news recipients" (p. 115). That is simply not the case. Both the data and the occasionally expressed opinion of the author indicate that individuals may be analyzed only insofar as they "are free to pick and choose among news offerings" (p. 115). Social persons perceive and interpret in socially patterned ways; we ignore this truth at our peril. Graber's exceptional data are not matched by an equally impressive analytical treatment of what the data might mean.

Networks, Exchange, and Coercion: The Elementary Theory and Its Applications. Edited by David Willer and Bo Anderson. New York: Elsevier–North Holland, 1981. Pp. xii + 240. $26.50.

Siamak Movahedi
University of Massachusetts, Boston

Networks, Exchange, and Coercion represents an interesting venture into the realm of theory construction in social relations. The goal of the book's authors is to construct a theory—based on the exchange model—that can explain human interaction on both micro- and macrolevels. The authors'

ambition shows in their attempt to incorporate enough of Marx and Weber in their discussion to appeal to sociologists in either camp.

The book consists of 13 papers with an introduction and an epilogue written by the editors. There is also a glossary of the terms they use. David Willer does most of the theoretical work in this volume. His three papers lay down the conceptual and formal foundation of what the various authors refer to as "the elementary theory of human interaction." Among the other papers there are some excellent pieces of ethnographic and cross-cultural case studies on the dynamics of exchange relationship and mechanisms of normative control in social, economic, and political spheres of life. I should add that, although the authors of the latter papers make occasional references to "the elementary theory," their works are actually independent and can be read without knowledge of "the elementary theory." However, they are all based on the general exchange model.

Steven Gilham's paper, "State, Law and Modern Economic Exchange," challenges the implicit assumptions of many economic and exchange theories and analyzes the theoretical functions of law, order, legal coercion, and the institution of property, the preconditions usually ignored or taken for granted. In general, this book is a valuable addition to the literature on exchange theory and economic models of human behavior. The authors show a great deal of awareness of the conceptual issues and problems of traditional exchange theories. Because much of this book is preoccupied with formalization or axiomatization of a structural theory of social interaction, I should like to confine most of my commentary to an analysis of this theoretical treatise.

The book begins with a discussion of the need for a *fully general* theory of human behavior, one which could take into account the biophysical, social, and cognitive aspects of human events. David Willer and Bo Anderson examine the works of Marx and Weber in regard to these three aspects. They argue rightfully that neither man had devised a fully general theory. Willer and Anderson then accept the challenge to develop such a theory. Further, they argue that "neither Marx or Weber followed through to develop a systematic scientific theory of social phenomena to be used in their substantive studies. That they did not was due to a crucial missing part in both theories, *a modeling procedure*. In fact, no social theory up until now has developed a modeling procedure. That will be the first task of the elementary theory" (p. 7).

After a critique of the "reductionistic operant exchange" theories of B. F. Skinner, George Homans, and Peter Blau, the authors expound their own theory—which allegedly explains human interaction and the rate of exchange in terms of the structural contingencies and social network rather than by the individual characteristics or reinforcement histories of the two actors involved in a relationship. They call this theory

the elementary theory of social interaction. This elementary theory is, according to the authors, based on a new conception of science, a conception that states that the core of any scientific theory is "patterned mental images." The patterned mental images of the external world are called models, and the procedure for forming these models is termed *modeling procedure*. The modeling procedure of the elementary theory is said to be that of interpreted network geometry.

The theory begins with some basic terms such as "sanction," "actor," and "potential network," and some derived terms like "exchange," "conflict," "coercion," and "system state." For instance, "conflict" is a dyadic social relationship where the sanction transmitted and the sanction received are both negative.

In addition to a number of definitions, the theory has two principles and two laws. The first principle is the familiar proposition of all economic models, that is, "All social actors act to maximize their expected preference state alteration [gain]" (p. 28). This statement is taken to be axiomatic and, as such, nonfalsifiable. Space limitation precludes a presentation of the two laws and the second principle of the theory. However, it is critical to point out that the other principle and the two laws of the theory are all analytic statements and, as such, tautological. Willer is fully aware of this point and notes that "the first principle of the elementary theory is a theoretic tool and cannot be falsified, for its referent is not any empirical event but the abstract objects and events of the theory. Similar statements could be made concerning the second principle and the two laws of the theory" (p. 29). The fully general theory of human interaction thus turns out to be nothing but a set of definitions or tautological statements. There is little doubt that, as the authors put it, this theory is based on a "new" conception of science. Given the principles, laws, and definitions of the elementary theory, the behavior of actors—transparent, abstract objects who have full and accurate system-relevant information—can be predicted with certainty. In fact, according to Willer, there is no room for probability in this theory.

I should admit that I have a great deal of difficulty understanding the value of the elementary theory. To begin with, I do not believe that this is a new *fully general* sociological theory of any sort. It is at best an elementary game-theoretic formalization of some simple interactional episodes. Of course, I may be wrong. In fact, Willer does predict the epistemological dilemmas of social scientists like me. He writes: "The use of abstract theory requires mental habits that are different from those typical of much mainstream sociology, political science and anthropology. To those accustomed only to generalization and other empirical modes of thought, mastering the theory will require a radical shift of perspective. (Two other groups, however, may be immediately comfortable with the

elementary theory: classical theorists, particularly those concerned with Marx and Weber, and formalists)" (p. 32). It is possible that, one day, I may change my mental habits, but until then I remain fully unconvinced about the value of such formalizations at this stage of sociological theory. I do not believe that the challenge in sociology today is one of formalization, axiomatization, or modeling. I am certain that, if Marx and Weber had theorized in terms of graphs, arrows, arcs, little circles, squares, and dots, we would not have heard of them today.

PHILOSOPHY AND SOCIAL CRITICISM
an international journal
editor: *David M.Rasmussen,* Boston College

Recent and Forthcoming Articles:

Agnes Heller:
"Rationality and Democracy"

Karl-Otto Apel:
"Hermeneutic Philosophy of Understanding as a Heuristic Horizon for Displaying the Problem-Dimension of Analytic Philosophy of Meaning"

Ferruccio Rossi-Landi:
"On Linguistic Money"

David M. Rasmussen:
"Communicative Action and Philosophy: Reflections on Habermas' *Theorie des Kommunikatuen Handelns*"

Umberto Eco:
"The Sign Revisited"

In modern industrial society reason cannot be separated from practical life. At their interface a critical attitude is forged. **Philosophy and Social Criticism** wishes to foster this attitude through the publication of essays in philosophy and

politics, philosophy and social theory, socio-economic thought, critique of science, theory and praxis. We provide a forum for open scholarly discussion of these issues from a critical-historical point of view.

Community Review

Journal for College Faculty in Two-Year Degree Programs

Community Review serves members of the two-year urban oriented community by providing a forum for the exchange of professional concerns and interests. A wide range of topics includes specialized sections on experimental programs, curriculum, population, conference news, and book reviews. *Community Review* encourages colleagues to share important research and articles which inform as well as demonstrate their authors' expertise.

Volume IV, Number 2, Spring 1982 features

DYNAMIC NEW PERSPECTIVES ON SOCIOLOGICAL INQUIRY

...FROM PRAEGER

GANG DELINQUENCY IN AN AMERICAN SUBURB
LOSERS IN PARADISE
Gene Muehlbauer and Laura Dodder

In describing the origin of a suburban "gang" whose members call themselves, ironically, the Losers, the book explains how, in the midst of a community which is modeled after the original American dream, class conflict and cultural alienation lead to the nearly complete disaffection of a large segment of youth from their peers, parents and established authority. *It provides insight into the genesis of individual and group deviance in suburban schools, and describes the toll which this process takes on individuals and the community.*

The book is organized in such a way as to aid the reader in understanding the social processes involved in the evolution of the Losers gang. To this end, most chapters find theory combined with description. When appropriate, the findings of other researchers about similar subject matter are included.
140 pp. 1983 $22.95
ISBN0-03-060313-7

CRISIS MANAGEMENT IN AMERICAN HIGHER EDUCATION
1960 TO THE PRESENT
Barbara A. Scott

A timely analysis of the problems facing American higher education, viewing these as symptomatic of deeper problems facing American society. Demonstrates that the currently popular managerial/administrative approaches to solving the problems—cutting costs, increasing efficiency, and clarifying lines of authority—ignore the broader "political" nature of the crisis. Provides historical background on the political development of today's crises, and presents a contemporary context for wide-ranging solutions.
ca. 300 pp. 1983 ca. $26.95
ISBN 0-03-062177-1

SENIOR SETTLERS
SOCIAL INTEGRATION IN RETIREMENT COMMUNITIES
Nancy J. Osgood

Presents a comparative study of social integration and community creation in three planned retirement communities. The communities included one all Jewish condominium, one mobile home community in Florida, and one affluent, middle-class community south of Tucson, Arizona. This study is grounded in a larger theoretical context of sociology. The dual concepts of community and social integration provide the theoretical framework in which the study is conceived and carried out.

In investigating the importance of age as the basis for development of community and social integration, the study examines the conditions under which older individuals living in a retirement community develop a sense of identity and a feeling that their needs are intertwined with others. Similarly, it addresses exactly when people in an age-segregated setting build community institutions and other forms of social organization to manage and regulate interaction within the community, and jointly confront problems they face in common.
ca. 256 pp. 1983 $27.95
ISBN 0-03-059792-7

ALSO OF INTEREST:
LIFE AFTER WORK
RETIREMENT, LEISURE, RECREATION, AND THE ELDERLY
edited by Nancy J. Osgood
384 pp. 1982 $29.95
ISBN 0-03-060437-0

WORK AND EQUALITY IN THE SOVIET UNION
THE DIVISION OF LABOR BY AGE, GENDER, AND NATIONALITY
Michael Paul Sacks
224 pp. 1982 $22.95
ISBN 0-03-046141-3

ENTREPRENEURSHIP
ITALIAN STYLE
Edith Kurzweil
ca. 208 pp. April 1983 ca. $23.95
ISBN 0-03-061709-X

PRAEGER Publishers
521 Fifth Avenue
New York, N.Y. 10175

CONTENTS

Review Essay

Book Reviews

IN THIS ISSUE

KAREN S. COOK is associate professor of sociology and director of the Center for Studies in Social Psychology at the University of Washington. She is currently engaged in experimental research on social exchange networks, focusing primarily on power and equity processes. She is also involved in field research on interorganizational relations within the hospital industry.

The late RICHARD M. EMERSON was professor of sociology at the University of Washington, where he was engaged in developing theory and conducting research on social exchange. In addition to laboratory work of the type reported in this issue, for the past decade he had also been involved in field-historical research on coercive state power in South Asia.

MARY R. GILLMORE received her Ph.D. in sociology from the University of Washington, where she is currently teaching sociology. Her research interests include the study of power, coalition formation, and structural correlates of social solidarity from a social exchange perspective.

TOSHIO YAMAGISHI received his Ph.D. in sociology from the University of Washington and is currently associate professor in the Division of Social Sciences at Hokkaido University in Japan. His research interests include social exchange, equity theory, models of information integration, and the sociology of knowledge.

DON PATINKIN, born in Chicago, received his Ph.D. from the University of Chicago. In 1949 he emigrated to Israel to join the staff of the Hebrew University of Jerusalem, where he serves as professor of economics. He is the author of *Money, Interest, and Prices* (1956, 1965), *The Israel Economy: The First Decade* (1959), *Studies in Monetary Economics* (1972), *Keynes' Monetary Thought: A Study of Its Development* (1976), *Essays On and In the Chicago Tradition* (1981), and *Anticipations of the General Theory? And Other Essays on Keynes* (1982) as well as numerous journal articles. He is a member of the Israel Academy of Sciences and Humanities and a foreign honorary member of the American Academy of Arts and Sciences.

JOHN A. LOGAN directs the programming of a multidisciplinary, environmental impact model at URS-Berger, a private firm. He is also developing a multivariate technique for the study of mobility under technological and social-structural constraints, a technique which complements that described in the present article.

FRED C. PAMPEL is associate professor of sociology at the University of Iowa. He is currently studying economic development and life-cycle patterns of labor force participation of males and females. He is also investigating the determinants of cross-national variations in governmental support programs for retirees and the aged. He has recently published a research monograph, *Social Change and the Aged,* that examines change in the labor force, income, and family status of the aged in the United States.

JANE A. WEISS, who died in 1981 after this research was launched, was assistant professor of sociology at the University of Iowa. Her interests included the political status of women, labor force participation, and income inequality, all studied from a cross-national, world-systems perspective. Her most recent work includes an article in the *American Sociological Review* (with Howard Aldrich), "Differentiation within the United States Capitalist Class," and a chapter on the political incorporation of women (with Francisco Ramirez) in *National Development and the World System* (edited by John W. Meyer and Michael T. Hannan).

MARYLEE C. TAYLOR is assistant professor of sociology at the Pennsylvania State University. Her primary research interest is the structure of white racial attitudes. Other current work focuses on the implementation of affirmative action and on methodological issues in the study of psychological androgyny.

NORMAN K. DENZIN is professor of sociology and humanities at the University of Illinois at Urbana-Champaign. He has recently completed a monograph on emotion, entitled *On Understanding Emotion*. He is at work on a series of studies that will be gathered together under the title *The Structures of Interpretation*.

PATRICK D. NOLAN is assistant professor of sociology at the University of South Carolina (on leave as a visiting associate professor of sociology at the University of North Carolina at Chapel Hill, 1983–84). He is interested in theories of macrosocial structure and change. He is investigating the effects of techno-economic heritage on Third World socio-economic growth with Gerhard Lenski and alternative theories of fertility stability and change with Ralph B. White.

MICHAEL P. MASSAGLI is a doctoral candidate in sociology at the University of Wisconsin—Madison. His current research is on trends in social mobility and fertility in the United States.

ROBERT M. HAUSER is Stauffer Professor of Sociology at the University of Wisconsin—Madison. His latest book, edited with D. L. Mechanic, A. O. Haller, and T. S. Hauser, is *Social Structure and Behavior: Essays in Honor of William H. Sewell*. In his current research, Professor Hauser is using data for primary respondents and their siblings in the Wisconsin Longitudinal Study to develop structural equation models of the stratification process with errors in variables and common family variance components. In the near future, he hopes to extend these analyses to comparisons of samples within the United States and to comparisons with similar data for other nations.

Information for Contributors

Editorial Procedures: All papers deemed appropriate for the *AJS* are sent out anonymously to readers. The editors rely heavily on the judgments of those readers, although they are not bound by them. To protect anonymity, only the title should appear on the manuscript. Attach a cover page with the title of the manuscript, your name, and your affiliation. All identifying references and footnotes should appear on a separate page. Manuscripts are accepted for publication subject to nonsubstantive editing with the understanding that the *AJS* has the right of first publication. Submission of a manuscript to the *AJS* is taken to indicate the author's commitment to publish in this *Journal*. No paper known to be under consideration by any other journal will be reviewed by the *AJS*. Upon publication, all rights, including subsidiary rights, are owned by the University of Chicago Press. The author retains the right to use his article without charge in any book of which he is the author or editor after it has appeared in the *Journal*.

Manuscript Acceptance Policy: While it is our policy to require the assignment of copyright on most journal articles and review essays, we do not usually request assignment of copyright for other contributions. Although the copyright to such a contribution may remain with the author, it is understood that, in return for publication, the journal has the non-exclusive right to publish the contribution and the continuing right, without limit, to include the contribution as part of any reprinting of the issue and/or volume of the journal in which the contribution first appeared by any means and in any format, including computer-assisted storage and readout, in which the issue and/or volume may be reproduced by the publisher or by its licensed agencies.

Preparation of Copy

1. Type *all* copy—including indented matter, footnotes, and references—double-spaced, allowing generous margins at top, bottom, and sides of page. Please do not justify right-hand margins. Number all footnotes, including acknowledgment footnote.

2. Type each table on a separate page. Refer to each table in numerical order in the text. Prepare tables without vertical lines.

3. Clarify all symbols with words in the margin of the manuscript. Encircle these and other explanatory notes not intended for printing.

4. Draw figures on white paper with india ink. Original or glossy print of each figure will be required if manuscript is accepted.

5. Include an abstract of not more than 100 words summarizing the findings.

6. *Four copies* of the manuscript and abstract, on white paper, must be submitted. Manuscripts should be addressed to the Editor of the AMERICAN JOURNAL OF SOCIOLOGY, 1130 East 59th Street, The University of Chicago, Chicago, Illinois 60637. *Manuscripts are not returned to authors.* Enclose

a self-addressed stamped postcard, containing the manuscript title, to obtain acknowledgment of manuscript receipt. No other acknowledgment will be sent.

7. Each article or research note submitted to the *AJS* must be accompanied by a check or money order for $10.00, payable to The University of Chicago Press in U.S. currency or its equivalent by postal or express money order or bank draft. Papers will not be reviewed until this fee has been paid. The submission fee will be waived for papers solely by student authors. Submissions from students should be accompanied by a faculty member's certification of their student status. Citizens of countries with restrictions on the export of U.S. dollars may request waiver of the submission fee.

Format of References in Text: All references to monographs, articles, and statistical sources are to be identified at the appropriate point in the text by last name of author, year of publication, and pagination where appropriate, all within parentheses. Footnotes are to be used only for substantive observations. Specify subsequent citations of the same source in the same way as the first one; do not use "ibid.," "op. cit.," or "loc. cit."

1. When author's name is in the text: Duncan (1959). When author's name is not in text: (Gouldner 1963).

2. Pagination follows year of publication: (Lipset 1964, pp. 61–65).

3. For more than three authors, use "et al." For institutional authorship, supply minimum identification from the beginning of the complete citation: (U.S. Bureau of Census 1963, p. 117).

4. With more than one reference to an author in the same year, distinguish them by use of letters (*a, b*) attached to the year of publication: (1965*a*).

5. Enclose a series of references with a single pair of parentheses, separated by semicolons.

Format of References: List all items alphabetically by author (providing the full list of multiple authors) and, within author(s), by year of publication, in an appendix titled "References." Examples follow:

Davis, K. 1963*a*. "Social Demography." Pp. 124–37 in *The Behavioral Sciences Today*, edited by Bernard Berelson. New York: Basic.

———. 1963*b*. "The Theory of Change and Response in Modern Demographic History." *Population Index* 29:345–66.

Goode, W. J. 1967. "The Protection of the Inept." *American Sociological Review* 32:5–19.

Moore, Wilbert E., and Arnold S. Feldman. 1960. *Labor Commitment and Social Change in Developing Areas*. New York: Social Science Research Council.

Sanford, Nevitt, ed. 1962. *The American College*. New York: Wiley.

Weber, Max. (1921) 1968. "Society's Problems." Pp. 12–16 in *Economy and Society*, edited by Guenther Roth and Claus Wittich. New York: Bedminster.

The Distribution of Power in Exchange Networks: Theory and Experimental Results[1]

Karen S. Cook, Richard M. Emerson, Mary R. Gillmore,
and Toshio Yamagishi
University of Washington

This paper presents a theoretical analysis of the structural determinants of power in exchange networks, along with research findings from laboratory experiments and a computer simulation of bargaining in network structures. Two theoretical traditions are dealt with: (1) point centrality in graph-theoretic representations of structure, as an approach to power distributions; and (2) power-dependence principles applied to exchange networks. Measures of centrality available in the literature have the advantage of being easily applied to large and complex networks. In contrast, power-dependence concepts were conceived for use in microsociology and are found to be cumbersome in the analysis of complex networks. But despite the relative difficulty of applying power-dependence theory to network structures, that approach generates hypotheses about power distributions which are confirmed at nearly every point in a laboratory experiment with five-person networks and at every point in a computer simulation of networks too large for laboratory study. In contrast, centrality measures applied to the type of networks studied fail to predict power distributions. Although centrality measures might predict power in some networks, their generality is limited. Toward resolution of the issues raised, this study offers two theoretical points: (1) a distinction between two different principles of "connection" in social networks suggests that current measures of centrality might predict power in one type of network but not in the other; and (2) it offers a first step toward a fusion of power-dependence theory and structural centrality in a way which might be general across networks of both types.

A review of the literature in sociology and anthropology over the past 15 years would show a virtual explosion in research dealing with social networks. This explosion is partially a result of the rapid improvement during this time period in the methodology for analyzing network data

(see, e.g., White, Boorman, and Breiger 1976; White and Breiger 1975; Burt 1976). Theoretical developments, however, have lagged behind methodological advances, so much so that Granovetter (1979) was prompted to caution that a "theory gap" exists. One reason for this state of affairs is that network methods have not been tied closely to existing bodies of sociological theory. More commonly, network properties have been examined descriptively (e.g., in detailed analyses of particular individuals' personal networks) or have been treated as variables to be "added to" other sets of factors in the explanation of some particular behavior (e.g., voting behavior, health-seeking behavior). Even the sophisticated techniques developed to render complex data sets analyzable (i.e., blockmodeling techniques) give us primarily new methods for detecting social structure, not a "theory of social structure."

Furthermore, the devices we use to represent networks—such as points, lines, edges, and geodesics—and the concepts we use to describe network properties—such as density, centrality, and degree of connectedness—are devoid of specific substantive meaning. A point can be a person, an organization, or any other entity, and a line can represent anything that can occur between two points (e.g., a friendship link, a business transaction, the flow of information, influence, resources, or energy). This abstractness has facilitated formal mathematical analysis, for example, the development of graph theory (Harary, Norman, and Cartwright 1965); but, as Foster (1979), Freeman (1979), and others have suggested, it has frequently presented problems in the interpretability of findings.

In this article we attempt to articulate theory concerning exchange networks with structural concepts drawn from recent work on social networks. Stimulated by Freeman's (1979) thoughtful essay, we focus on centrality and its relation to power in networks of connected exchange relations. Using important theoretical distinctions taken from exchange theory (Emerson 1972), we demonstrate that in certain types of exchange networks existing measures of structural centrality (as identified by Freeman 1979) have limited utility in predicting the locus of network power. Empirical work of this type can help not only to clarify the link between centrality and power but also to identify the limitations of existing measures of centrality and to suggest fruitful directions for the development of more general theory. First, we introduce relevant theoretical notions; then we describe empirical results and present conclusions concerning centrality and the distribution of power in exchange networks.

EXCHANGE NETWORKS: BASIC CONCEPTS

Many of the social networks of interest to social scientists can be analyzed fruitfully as exchange networks, provided that the specific content of the

social relations in the network involves the transfer of valued items (i.e., the provision of information, affection or approval, advice, or more tangible things like goods and direct services; see Sarason et al. [1978] on this point). The concept "exchange network" has the theoretical advantage of allowing the extension and application of already well-developed dyadic conceptions of exchange (e.g., Homans 1961, 1974; Emerson 1962, 1972) to more macro, N-actor levels of analysis.

On the basis of Emerson's (1972) earlier work, an exchange network can be defined as consisting of: (1) a set of actors (either natural persons or corporate groups), (2) a distribution of valued resources among those actors, (3) for each actor a set of exchange opportunities with other actors in the network, (4) a set of historically developed and utilized exchange opportunities called exchange relations,[2] and (5) a set of network connections linking exchange relations into a single network structure. Thus an "exchange network" is a specific social structure formed by two or more connected exchange relations between actors, with "connection" defined as follows:

> *Definition 1:* Two exchange relations between actors A-B and actors A-C are connected to form the minimal network B-A-C to the degree that exchange in one relation is contingent on exchange (or nonexchange) in the other relation. (*a*) The connection is positive if exchange in one relation is contingent on exchange in the other. (*b*) The connection is negative if exchange in one relation is contingent on nonexchange in the other.

The importance of an explicit theoretical treatment of the concept "connection" in any theory of social networks has not been fully recognized. That any two dyads, A-B and A-C, have one member in common (i.e., actor A) does not necessarily imply that these two relations are connected and thus represent a three-actor network. As a result, for exchange networks, common membership is not sufficient as a "connecting principle." If the A-B and A-C exchange relations are connected (by definition 1), they define the minimal network or component, B-A-C, of possibly larger structures, each such component being either positively or negatively connected within itself. Larger network structures might therefore consist of purely positive connections, purely negative connections, or some mixture of both types.

The distinction between negative and positive connections has significant theoretical implications for network analysis which have yet to be fully developed. For example, the discussion thus far has said nothing about the social conditions producing one or the other type of connection.

---

[2] The set of exchange relations is properly viewed as a subset of exchange opportunities. Members of an exchange relation are assumed to have some degree of "commitment" to the relation, relative to other potential alternatives (see Cook and Emerson [1978] regarding commitment).

Generally speaking, one expects that if B and C are alternative exchange partners for A, in the sense that B and C are substitutable as sources, then the connection is negative.[3] Alternative sources thus introduce an element of negativity or competition (see Cook 1978) into the exchange system. Dating networks and friendship networks are typically negatively connected throughout. On the other hand, if a resource obtained from B is required by A for interaction with C (e.g., as when A is a broker), the connection at A is positive. Large networks completely positive in form are probably very rare because of the frequent existence of alternative sources. "Mixed" structures, we suspect, are much more common. For example, exchange in the Kula ring, described by Malinowski (1922), involves a specific complex pattern including exchange connections of both types (see Emerson 1981). Similarly, larger networks involving brokers typically entail both positively and negatively connected exchange relations. Marsden's (1982) simulation of "brokerage" is a study of "mixed" networks.[4] For both substantive and analytical reasons, we begin our investigation with a focus on purely negatively connected networks.[5] However, research concerning purely positively connected networks and mixed networks is currently under way.

The concept "network connection" allows us to specify the boundaries of concrete networks and develop a theory in which events happening at any location in the network have predictable repercussions within the boundaries of the network. The concept of connection and the distinction between two basic types[6] is one of the primary features distinguishing the

[3] In addition to specifying the conditions under which different types of connection are likely to emerge, it is theoretically interesting to specify the mechanisms which alter the nature of the connection. For example, a negative connection might be transformed into a positive connection through "product differentiation" or some other type of resource value differentiation. Well-known examples of these types of processes exist in economics (e.g., product differentiation) and sociology (e.g., division of labor or role differentiation).

[4] While Marsden's (1981a, 1981b, 1982) work is based on Coleman's model of exchange, the particular interest and control structures he specifies in his simulation (Marsden 1981b) combine to produce elements of both positive and negative connection in the networks he investigates.

[5] The exact character of the connections forming natural social networks must be determined by research directed to that end. That is our major point in introducing the concept. Given the difficulties such field research can encounter, it is essential to develop theory regarding network connections. Laboratory research, in which networks connected in known ways can be created and studied, has much to offer in the development of such theory. That the networks established in our laboratory resemble in exact detail no "real" network should neither surprise nor bother the reader. External validity is, of course, an important concern. This means that definitions and propositions which can be applied both inside and outside the laboratory must be developed through our efforts at theory construction. We have reason to believe that, when that is done, laboratory work can inform field research.

[6] In addition to the sign of the contingency (positive vs. negative) linking exchange relations, it is important to note that network connections can vary in strength. We will not develop a quantitative concept of connection in this article.

exchange approach to network structures from other theories or methods of network analysis. While exchange networks have empirically determined boundaries, often the actors themselves are not aware of those boundaries. For example, in a three-actor network, B-A-C, actors B and C might not even know of one another. Similarly, A might have no knowledge about possible network relations beyond or between B and C. This has important implications for the analysis of social structure. Frequently there are no consensually defined network boundaries, even though boundaries do exist. Thus, participation in a network typically is not based on "membership status." Instead, actors can be viewed as relatively autonomous decision makers occupying "positions" in a structure which frequently extends beyond their own awareness.

Position in an Exchange Network

For simplicity, an exchange network can be represented as a digraph (see Harary et al. 1965; Berge 1962) or as a flow network (Harary et al. 1965; Busacker and Saatz 1965). Points in a network graph represent individual or corporate actors; lines or edges represent exchange relations. We use the notion of "residual graph" from graph theory to specify what we mean by position in a network. A residual graph (or matrix) is obtained by the removal of a specified point from a parent graph. The use of this concept allows us to locate sets of actors (identified as points in the graph) who have structurally similar locations in a network. We refer to these actors as occupants of the same position. Thus, we define position in graph theory terms as follows:

> *Definition 2:* A position in a graph or network is a set of one or more points whose residual graphs are isomorphic.

The concept "position" is important for two reasons: (*a*) it helps simplify the analysis of otherwise more complex networks, and (*b*) it has been demonstrated to be an important determinant of behavior in exchange networks (see Cook and Emerson 1978). Figure 1 portrays some of the network structures studied in our laboratory. Letters identify network positions, and numerical subscripts represent individual actors as occupants of each position.

In the networks shown in figure 1: (1) if each actor has a resource which the other actors value and each actor values all other actors' resources, (2) and if each line represents an opportunity to exchange these valued resources, (3) then the patterns of lines displayed in this figure can be considered exchange "opportunity structures." Any opportunity actually used will involve a mutually beneficial two-way transfer or exchange of resources. Within these opportunity structures, over time, networks of

connected exchange relations emerge. In our laboratory research the experimenter determines the opportunity structures shown in figure 1. The subjects in these network studies conduct exchanges within the constraints set by the opportunity structure, forming exchange relations and a network of exchange through their actions. In the particular structures diagramed in figure 1, solid lines represent more profitable exchange opportunities than broken lines (by a factor of three to one); thus some opportunities will be experienced as clearly more beneficial than others. The less favored opportunities should not lead to the formation of continuing exchange relations. Therefore, the actual networks expected to emerge are those represented by the solid lines in figure 1. These emergent networks are negatively connected. If an actor has two solid-line opportunities, the two partners represented are fully interchangeable as sources of benefit. With finite exchange time available, any use of one opportunity means that another opportunity is forgone.

An important feature of our laboratory research is that the actors located in the structure have no knowledge of the network beyond their own opportunity set. Thus in figure 1, positions A, C, D_2–D_4, and E_3–E_8 are all identical from the occupants' viewpoint. But, as positions in a network, they differ from each other in one respect—the nature of the remote structure they are embedded in, which transcends the occupants' knowledge. This feature allows us to examine "purely" structural deter-
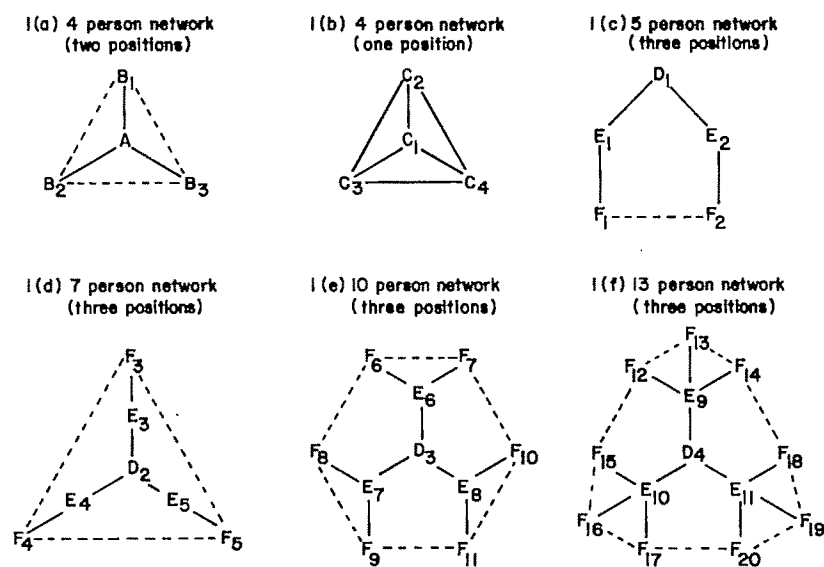

FIG. 1.—Exchange networks studied in laboratory experiments and computer simulations. (Positions are identified by letters, occupants by numerical subscripts. Lines represent the negotiated exchange of 24 points [solid lines] or eight points [dashed lines].)

minants of behavior. In particular, the distribution of power can be studied as a function of position in an opportunity structure. Furthermore, if we let the actual amount of resources exchanged in any given transaction be determined by negotiation, so that benefit, while mutual, is potentially unequal, power use can be measured in terms of the actual benefits obtained through negotiation.[7]

CENTRALITY AND POWER IN NEGATIVELY CONNECTED EXCHANGE NETWORKS

In a previous experiment (Cook and Emerson 1978) on the distribution of power in negatively connected exchange networks, the structures diagramed as figure 1a and b were studied. The results concerning power use demonstrated that power is a function of position in the network, even when the position occupants are ignorant of the actual network structure and their own position in it. More specifically, power was found to be concentrated in position A, the most central position, relatively absent in the peripheral position B (see fig. 1a), and evenly distributed across the occupants of position C (in the power-balanced network, fig. 1b).[8] The outcome of that experiment was predicted on the basis of simple power-dependence reasoning (i.e., A is less dependent than B and C_i and C_j are equally dependent). But the results could have been predicted parsimoniously on the basis of structural centrality, if power is hypothesized to be a function of centrality. Thus the question arises, Do predictions based on power-dependence notions and those based solely on structural centrality yield the same results in negatively connected networks? We turn to two bodies of theory and an experiment to provide an answer. The network diagramed in figure 1c is the one analyzed in the actual experiment. Simulation results are presented for networks 1d–f. In order to examine the generality of the link between centrality and power in negatively connected exchange networks, we moved to the analysis of larger, more complex networks in which not all actors have direct access to one another and no actor enjoys simple monopoly power (as A does in fig. 1a).

[7] In order to examine power use we removed from the situation factors known to inhibit the exercise of power. For example, actors in a network were not informed about the profits of other actors in the network. This procedure effectively removes the operation of equity concerns from the negotiations (see Cook and Emerson 1978).

[8] The comparison between these two networks can be considered an empirical comparison between a "star" network (fig. 1a) with maximum point centrality (i.e., one clearly dominant position) and an "all-channel" or completely connected network (fig. 1b) in which access is equalized across the network. Marsden (1981b) would refer to these two types of networks as "restricted access" vs. "open access" networks, respectively, with "access" referring to access to highly valued resources.

Point Centrality

Structural centrality is one of the most frequently discussed properties of networks.[9] Freeman (1979) identified three major conceptions of point centrality, each associated with a different underlying measurement approach: degree-based measures, betweenness measures, and closeness-based measures. A degree-based measure (see Nieminen 1974; Freeman 1979) is a count of the number of adjacencies for a point. As Freeman (1979, p. 221) notes in the context of communication networks, "The degree of a point is viewed as an index of its potential communication activity." Betweenness measures are based on the "frequency with which a point falls between pairs of other points on the shortest paths [or geodesics] connecting them" (Freeman 1979, p. 221). Points central in this respect exhibit potential for control, since according to Freeman (1979, p. 221), strategic location on paths linking pairs of points provides potential influence in communication networks through the "withholding or distorting of information in transition." Finally, closeness-based measures (e.g., Sabidussi 1966) give an index of the extent to which a particular point is "close" to all other points. This is a distance measure that counts the number of edges or lines in the paths (or geodesics) linking two points. Centrality in this case is indexed by the "shortest" distance score of one point to all others. Freeman (1979, p. 224) interprets this measure (or set of measures) as an indicator of the extent to which a point can "avoid the control potential of others." In the case of communication networks a "central" point, being close to other points, is less dependent on intermediaries for relaying information. Freeman concludes his article by stating that substantive concerns must determine the appropriateness of a particular conception and measure of centrality.

For some networks these three types of measures give the same results with respect to identification of the most central point. For example, all agree that the center of a "star" configuration is most central (e.g., point A in fig. 1a). However, in other types of networks there is marked departure from agreement in the assessment of point centrality. For example, even in the fairly simple networks in figure 1c–f, there is some variation in the predictions regarding point centrality. Closeness and betweenness measures consistently identify point D as the most central position. Thus an ordering of point or position centrality based on these two types of measures suggests: $D > E > F$ in networks 1c–f. Different orderings can

occur when a degree-based measure, counting only adjacencies, is used. Degree-based measures suggest the following orderings of positions with respect to centrality in these same networks: (1) for networks 1c and 1e, $D = E > F$; (2) for network 1d, $D > E > F$; (3) for network 1f, $E > D > F$.[10]

The major conceptual weakness of this measure is that it is highly localized: it takes into account only direct links, not indirect effects or paths. Thus the other two measures are superior for our purposes because they take into account the total structure of the network by focusing on characteristics of the paths or geodesics which link all pairs of points in the graph. Since our primary interest is in centrality as a network-wide indicator of power, we set aside degree-based measures as theoretically inadequate for the task. That leaves closeness and betweenness measures as potentially useful conceptions of centrality in exchange networks. The virtue of these measures is that, unlike the dyadic conception of power-dependence theory, these measures of centrality incorporate the entire structure into the centrality score assigned to each position. Thus centrality is a theoretical direction worthy of close scrutiny in experiments on more complex networks (e.g., fig. 1c–f).

Centrality has been defined as one of the most important characteristics of positions, not only in communication networks (see, e.g., Bavelas 1948, 1950; Leavitt 1951; Shaw 1954, 1964; Mackenzie 1966) but also in sociometric networks (see, e.g., Moreno 1934, 1943) and interorganizational networks (see, e.g., Rogers 1974; Miller 1980; Mizruchi and Bunting 1981). Thus it is logical to examine its role in exchange networks. Interest in point centrality has been fueled partially by the empirical demonstration that power and influence seem to be a function of the centrality of one's position in a social system (see Hopkins 1964). For example, Marsden and Laumann (1977, p. 217) state that "those persons at the center of the network, on whom the more peripheral actors are dependent, are the most powerful actors in the system." In other work (see Laumann and Pappi 1973, 1976), this finding is referred to as the principle of "integrative centrality." Marsden and Laumann (1977, p. 224) also note that "power as computed by the Coleman model reflects the relative centrality of an actor in a network of dependency relations." The parallel use of the terms "centrality" and "dependency" in the work cited is noteworthy. Their relation to one another is the basic focus of this theoretical section.

If power is indeed a function of centrality, measured in terms of closeness or betweenness, then we arrive at the following experimental hypothesis concerning the link between power and centrality in the negatively

---

[10] Centrality scores based on degree-, betweenness-, and closeness-based measures can be computed easily for each network position in fig. 1c (see Freeman [1979] for computations).

connected exchange network graphed in figure 1c, the object of the experiment reported here:

> *Hypothesis 1:* In the network portrayed in figure 1c, $D > E_i > F_j$ in power if either closeness or betweenness-based measures of point centrality are used.

This same prediction holds for the networks diagramed in figure 1d–f, tested in a series of simulation experiments reported below.

We do not offer hypothesis 1 with theoretically based confidence. Instead, we offer it because (a) structural principles are desirable in order to advance exchange network theory, (b) point centrality appears to be the best currently available candidate because of its relation to power, and (c) this hypothesis makes explicit the predictions which the best measures of point centrality would make, if applied to the networks under study here. Since we are dealing with negatively connected exchange networks (as in the previous experiment reported in Cook and Emerson [1978]), we are especially interested in the general applicability of centrality notions to predictions concerning the locus of power in this type of network. It is important to note that many centrality measures were originally developed to apply to networks in which resources (or bits of information) flow through intermediary points. In our negatively connected networks, however, the flow of resources is between two adjacent actors (i.e., points) with no intermediary; that is, there is a direct as opposed to an indirect exchange of resources. If point-centrality measures do not predict well in negatively connected networks, either centrality is not linked to power in such networks or these measures are limited to certain kinds of networks and we must specify more carefully the substantive meaning of existing measures of centrality. Freeman (1979) concurs. Such specification is necessary if we are to develop a more general theoretical conception of centrality which accommodates qualitatively different types of networks.

POWER AND DEPENDENCE

While the relation between power and centrality is intuitively compelling, it has not been given an explicit theoretical interpretation. As a step toward that end, we return to power-dependence theory. The following definitions are taken, with minor modifications, from Emerson (1962, 1972):

> *Definition 3:* In any dyadic exchange relation A_x; B_y (where A and B are actors, and x and y are resources introduced in exchange), the power of A over B (P_{AB}) is the potential of A to obtain favorable outcomes at B's expense.

> *Definition 4:* The dependence (D_{AB}) of A on B in a dyadic exchange relation

(e.g., A_x; B_y) is a joint function (1) varying directly with the value of y to A, and (2) varying inversely with the availability of y to A from alternate sources.

On the basis of Emerson's arguments (1962, 1972), we assert the following fundamental relationship between power and dependence: $P_{AB} = D_{BA}$.

Power-dependence concepts deal with the distribution of power between two partners in a dyadic exchange relation; thus they are not well suited for analyzing the power distribution among positions in an extended network. However, in very simple networks of the sort studied in the laboratory, power-dependence analysis can be applied, one relation at a time (and only one relation at a time), across the network. The results of this analysis yield three unambiguous predictions concerning the locus of power in these negatively connected networks. The predictions are as follows:

> *Hypothesis 2:* As the exchange process proceeds through time, the occupants of position E will display more power use than the occupants of positions F and D. This display of greater power use will take two forms: (a) an increase over time in the amount of benefits received from exchange at position E, and, as a result, (b) a greater absolute level of exchange benefit obtained by the occupant of position E by the final exchange phase.

> *Hypothesis 3:* The differential power use of E over F will be displayed before the power use of E over D (since the latter process is, in theory, predicted to be a result of E's power use over F).

> *Hypothesis 4:* In the final or stable phase of power use, the occupants of position E will exert equal levels of power over the occupants of positions F and D.

According to these hypotheses, the predicted ordering of positions with respect to power in these networks at equilibrium is $E > D = F$, which contradicts the prediction based on point-centrality measures (hypothesis 1).

The reasoning behind these predictions follows directly from power-dependence principles and the concept, exchange connection, when they are applied to all of the dyadic relations in the network. First, in negatively connected networks, any two lines joined at a point provide that point with "alternative sources" of value, as stated in definition 4. Therefore, if the relative value of resources is held constant (as in our experiment),[11] the structure of the network determines the relative dependencies throughout the network. Second, while positions D and E have equal access to

---

[11] By "constant" we simply mean that the value is assumed to be the same across the conditions in our experiment. This assumption is reasonable since, although it is obvious that values vary across individuals, if subjects are assigned randomly to network positions as well as experimental conditions, there should be no systematic differences by condition.

resources, each having two valuable partners, their partners do not have equal exchange opportunities. In figure 1c, unlike D, E_i has one very dependent partner over whom he or she has power and from whom, therefore, he or she will obtain high benefits assuming that E_i uses power in a "rational" way.[12] As a result, D's apparent power equality with E_i (based on an equal number of alternatives) will be short lived, for D must compete with F_j for access to E_i and F_j has no alternatives. Stated analytically, F-E-D is a negative connection or purely negative component, as are all the connections in these networks. Thus, D (the occupant of the central position) is ultimately as weak as all of the F_j (peripheral actors), with E_i emerging as the most powerful. These predictions hold even when none of the actors has knowledge about the network beyond his own immediate partnerships.

Reasoning from power-dependence theory can be carried still further. In definition 4, two variables govern dependence. One of these, the availability of valued resources, operates through position when network structures are involved, resulting in a social structural determinant of power. That is, position in the network determines availability. The other variable governing dependence, the relative value of the resources introduced at various positions, was held constant in the above predictions. If resource value is allowed to vary—as it does in nature—then it will confound the foregoing network structural determinants of power. However, if one varies resource value between networks while holding it constant (either high or low) within networks, still another hypothesis can be advanced.

If the incentive to exchange is high (because the resources exchanged are highly valued) throughout the network, then position will create differentials in resource availability; the latter determine dependence and thus power. Predictions 2, 3, and 4 assume some incentive to exchange. In contrast, if the incentive to engage in exchange is uniformly low, no actor in the network will be very dependent on the others and the potential power inequalities will be reduced across the network regardless of position. Therefore,

  Hypothesis 5: The effects implied in hypotheses 2, 3, and 4 will be more

[12] This assumption is necessary theoretically since it allows us to derive testable predictions concerning manifest power from principles dealing with potential power. In our experimental setting, by "rational" we mean that each actor in the network explores alternative sources of benefit in the network (a) through extending offers to others and (b) by comparing offers and counteroffers from others. Each actor maximizes benefit by (a) accepting the better of any two offers, (b) lowering offers when offers go unaccepted, and (c) holding out for better offers when it is possible to do so. This is clearly a testable assumption, but all one could conclude from evidence to the contrary is that sometimes subjects in our laboratory act "irrationally." We have examined empirically some of the conditions under which these conditions do not hold (e.g., when equity concerns are operative; see Cook and Emerson [1978]).

pronounced under conditions of high exchange incentive than under conditions of low exchange incentive.

POWER AS POTENTIAL, POWER USE, AND EQUILIBRIUM

Power, in definition 3, is conceived as a potential for gaining increased benefit at the other's expense in a dyadic exchange relation. In applying this conception of power to networks of negatively connected exchange relations, we have seen above that the structure helps determine that potential for each position in the network, but the occupants of these positions might use their power in varying degree or at a variable rate. If power is ever used to its theoretical limits, then, in our research setting, the less powerful actor could receive no more benefit from the more powerful actor than is obtainable from the best alternative source. This would define the theoretical equilibrium point.

Considering the structures in figure 1, if solid lines represent 24 units of profit to be divided through exchange and dashed lines represent only eight units of negotiable profit, then the best alternative source for all occupants of position F is fixed at four points (8/2). Thus, if power were exerted to its absolute maximum in these networks, all occupants of position E would obtain 20 points per exchange, while all occupants of positions F and D would obtain four points per exchange when equilibrium is reached (i.e., when the exchange ratios have stabilized). If there are restraints on the exercise of power (e.g., equity concerns or less than fully rational negotiation), equilibrium will be reached somewhere short of this maximally "exploitive" exchange ratio (see Cook and Emerson 1978).

But regardless of the particular equilibrium point reached over time in any specific network, the rate at which power use approaches the equilibrium level will be a function of the relative availability of resources to the actors in the network (i.e., their relative dependencies). This is a direct extension of power-dependence reasoning which can be investigated in our laboratory. Let us develop this reasoning and then derive hypotheses to be examined in a series of simulation experiments conducted on networks 1c–f.

What varies across these networks is not only the size of the network, but, more important, the relative availability over time of highly valued resources to the occupants of positions D and E_i. Relative availability of resources from alternate sources determines relative positional dependence (see definition 4). Relative positional dependence across the network of connected exchange relations determines power as evident in the chain of reasoning developed above (e.g., hypotheses 2–4). As a function of F_i's dependence on E_i, the relative availability to D of valued resources from

E declines over time (even though D and E have an equal number of alternate sources).[13] Thus D's dependence on E_i increases, with the result that E's power over D increases (since $P_{ED} = D_{DE}$). The outcome of this process is a reduction of power at the "center" in these networks. This chain of reasoning based on the rather cumbersome application of dyadic power-dependence notions and the concept of negative connection yields fairly clear predictions concerning the rate at which equilibrium is reached in the distribution of power in these exchange networks:

> *Hypothesis 6:* E's use of power over D will emerge more slowly in network 1*d* than in network 1*c*.

This prediction is based on the fact that the increase in the number of alternatives for D in network 1*d* increases the availability of resources to D and thus may result in an initial power advantage for D over E, but the advantage will be eroded over time as predicted above, because of the decreasing availability of valued resources from E. This decrease should occur later in network 1*d* than in 1*c*, retarding somewhat E's rise to power in 1*d*. Furthermore, it can be predicted that:

> *Hypothesis 7:* E's use of power over F will emerge more quickly in network 1*f* than in 1*e*, where it will emerge more quickly than in 1*d*.

This prediction results from the decrease in dependence of E_i on F_i across these three networks. The dependence of F on E_i is not altered across these networks, but E gains power over F to the extent that the resources F_i has to offer are available from other sources (i.e., other F_j). In addition, as stated in hypothesis 3 above, E_i will display a power advantage over D based indirectly on E_i's power over F. Thus it follows from hypotheses 3 and 7 that, since E's power over F is emerging more rapidly across these networks, E's power over D should also emerge more rapidly. Therefore:

> *Hypothesis 8:* E's use of power over D will emerge more rapidly in network 1*f* than in 1*e*, where it will emerge more rapidly than in 1*d*.

An important point to be made is that treating number of alternatives as a perfect indicator of resource availability can result in erroneous predictions when applied to connected sets of exchange relations and suffers from the same deficiency as a simple degree-based measure of

---

[13] Various theorists (see Blalock and Wilken 1981; Marsden 1982) treat number of alternatives as the main determinant of dependence (along with resource value). While this may seem to be implied by Emerson's (1962, 1972) definition of dependence, it is important to note that the operative term is "resource availability," which only under certain conditions translates directly into number of alternatives. As one anonymous reviewer cleverly stated, "It makes no difference how many bad sources of supply a position has." Thus it is important to distinguish theoretically between resource availability and the number of resource suppliers.

centrality. In theory, availability and number of alternatives must be kept analytically separate even though under certain conditions number of alternatives may completely determine availability (e.g., as in the case of a network in which there are no indirect paths).[14] Thus, in network 1d, for example, position E is predicted to emerge as more powerful than D over time, despite the fact that D has access to more alternative exchange partners and appears to occupy a more central location in the network.

THEORETICAL STRENGTHS AND DEFICIENCIES

Before we turn to experimental results, deficiencies in both of the theoretical approaches we have explicated should be noted. The issue before us in the following experiment is not which approach makes correct predictions in this case but, rather, how best to integrate network-structural principles and power-dependence theory to explain the dynamics of power in exchange networks. The difficulty with power-dependence concepts, as they now stand (e.g., Emerson 1962, 1972), is that they are too closely bound to dyadic analysis. The reasoning behind hypotheses 2, 3, and 4 is complex as a result. Yet, within this limitation, power-dependence theory has the virtue of being closely coordinated with concrete behavioral concepts and observations. Furthermore, it offers an intuitively appealing theoretical interpretation of the notion of centrality. Marsden and Laumann (1977), and others, as noted above, have attempted to relate dependency notions to power and centrality in networks. Power-dependency theory may help accomplish this task.

In contrast, the approach to power through point centrality of positions has the virtue of taking the structure of an entire network into account in specifying at once a degree of centrality (and thus a power level) for every position in that structure. Because of the formal mathematical properties of networks, such analysis can be applied to very complex structures. But this approach is weak where the other one is strong. The link between centrality and power is largely intuitive; and the abstract graph-theoretic networks to which these centrality measures have been applied are only loosely coordinated with the social interactive networks they represent. For example, the concept of connection, which is so fundamental in our substantive theory of networks, to our knowledge has not been incorporated in any formal network model. As a result, previous approaches cannot make differential predictions concerning the locus of power in positively and negatively connected networks. Thus we did not

[14] For example, in a star network all peripheral actors have direct access to one and only one source of valued resources. Adding more peripheral actors to the network simply increases the number of suppliers to the central actor; it does not alter the relative dependence of those on the periphery (unless there is a very finite supply of resources at the center).

design the following study as a "critical" test between two bodies of theory. Instead, we hoped to gain an empirical base for further theoretical development which would facilitate the integration of these two research traditions.

THE EXPERIMENT

An experiment was designed to test the foregoing predictions derived from point-centrality and power-dependence notions. The experiment was conducted in a computerized laboratory, using methods described more fully in Cook and Emerson (1977, 1978).

Briefly, subjects were recruited from undergraduate classes and campus newspaper ads. Emphasis was placed on the desire to earn money as a motive for taking part in the experiment. After a brief collective orientation, each subject was taken to a private room containing a computer terminal. All terminals are joined to a minicomputer in the laboratory which is programmed to allow certain terminals to communicate with certain other terminals. This procedure gives the experimenter control of the network of exchange opportunities.

Within the opportunity structure set by the experimenter, subjects negotiated with one another for "profit points" by sending offers and counteroffers until trade agreements were reached. Each transaction involved the division of a constant sum of points (either 24 or eight as shown in fig. 1) between two bargaining partners in that transaction. However, the subjects did not know the constant sum and therefore could not compare their own with the other's benefits. In this way principles of "equity" were effectively prevented from operating in this laboratory setting.

The total time of 81 minutes spent in the exchange process was divided into 27 transaction periods of 180 seconds each. Each person was allowed to complete only one transaction per period. Therefore, the network created in the laboratory was negatively connected: exchange in one relation was contingent on nonexchange in other relations during a given transaction period. That is, the use of one exchange opportunity meant that other opportunities had to be forgone during that time period.

Design Features

The experimental design involved the following features:

Network structure.—The network shown in figure 1c was used. Five persons in one network were treated as one experimental case or data unit.

Measurement of the dependent variable, power use.—In each exchange relation in the network (shown as a solid line in fig. 1), two persons

290

negotiated over the division of 24 points convertible into dollars. The use of power of one person relative to the other is measured as the number of points obtained through negotiation.

Incentive manipulation.—The manipulation of exchange incentive was straightforward. Both the amount of fixed wages the subjects would receive during the experimental session and the value of the profit points they could obtain by completing trade agreements were varied. These conditions were operationalized as follows: (1) In the high-incentive condition, subjects were paid $0.25 per hour while each profit point obtained through exchange was worth 2.5¢. As a result, most of their pay was derived through exchange. (2) In the low-incentive condition, the fixed wage for participation was $3.00 per hour and the value of each point was 0.5¢. Subjects in this condition therefore derived most of their pay through a fixed wage and were less dependent on making exchanges in order to derive pay.

Design.—The design of the experiment was a factorial type containing two between-subjects variables: (1) exchange incentive (high vs. low), and (2) subject gender (male vs. female). There was one within-subjects variable: trial blocks (there were 27 trials aggregated into three trial blocks each containing nine trials). Within each sex, subjects were randomly assigned to experimental conditions including positions within networks. Five cases were included in each cell of the design (a case is a five-person group).[15]

Subjects.—A total of 100 university students (50 male and 50 female) took part in the study described as an investigation of negotiated trade agreements.

RESULTS

Manipulation Check

As a partial check on the exchange incentive manipulation, subjects were asked on the postexperimental questionnaire to indicate how important earning money through exchange was for them. The results of a two-way (sex × incentive) analysis of variance on this item indicated a significant main effect for incentive ($F = 8.17$, $df = 1,96$, $P < .01$). The means ($\bar{X}$)

---

[15] Eight subjects were scheduled per session to insure that the subjects would not discover the identity of their bargaining partners. The extra three subjects, randomly selected from the eight, participated in a three-person replication experiment. The orientations for both experiments were identical and the computerized system allowed us to run more than one experiment simultaneously. The subjects knew only that they would have two exchange partners randomly selected from the seven others present and were not aware that two experiments were being run. Thus, they did not know the exact size of the networks or the nature of the exchange connections among the remaining participants.

indicated that the importance of earning money through exchange was greater for subjects in the high-incentive conditions than for those in the low-incentive conditons ($\bar{X} = 5.42$ vs. $\bar{X} = 4.52$ on a seven-point scale where 7 represents extremely important). No other effects were significant.

Power and Network Position

If power is indeed a function of point centrality, then, assuming actors use their power, the occupants of position D should evidence more power use than the occupants of the more peripheral positions, E and F. However, on the basis of power-dependence principles, we predicted in hypothesis 2 that occupants of position E would emerge as most powerful. Hypothesis 2 implies two findings: (a) a systematic increase over time in the amount of profit E is able to obtain in exchanges with both D and F, and (b) ability on the part of E to obtain better than half the total profit available per dyadic exchange with both D and F (i.e., E should be able to obtain significantly more than 12 points per exchange on the average since there are 24 points available for each exchange involving E).[16] Furthermore, hypothesis 5 predicts that these power differences will be more pronounced under conditions of high exchange incentive than under conditions of low incentive.

The profit data for E's exchanges with both D and F in network 1c under conditions of high and low exchange incentive are displayed in table 1. To test the hypothesis that E's profits from exchanges would increase over time, separate three-way analyses of variance (incentive × sex × trial block) for designs containing a repeated measure were performed on the profit obtained by E in exchanges with both D and F. As implied by hypothesis 2a, a significant main effect for trial blocks was obtained for the former (E's exchanges with D: $F = 3.34$, $df = 2,32$, $P < .05$) as well as the latter (E's exchanges with F: $F = 5.89$, $df = 2,32$, $P < .05$). Inspection of the cell means in table 1 indicates that these effects were due to an increase in the profit received by E over time as predicted by hypothesis 2a.

Hypothesis 5 implies that this increase in E's profits over time will be greater under conditions of high exchange incentive than low incentive.

---

[16] Since we are interested primarily in the effect of network position on the exercise of power, no distinction is made in our analyses between exchanges involving different occupants of the same position. The data for occupants of identical positions were averaged. For example, the profit obtained by E in exchanges with D was calculated as the average profit obtained by E_1 and E_2 in all dyadic exchanges with D. Similarly, the profit obtained by E in exchanges with F was calculated as the average profit of E_1 and E_2 obtained in all dyadic exchanges with F_1 and F_2.

In other words, the interaction of incentive and trial blocks should be significant. For E's exchanges with D, this interaction effect was significant ($F = 3.69$, $df = 2,32$, $P < .05$). Inspection of the trial means in table 1 indicates that this effect was due to the differential rate of profit increase over time in the two incentive conditions as predicted by hypothesis 5. For E's exchanges with F, however, the interaction was not significant.[17] Thus hypothesis 5 received support in the case of E's exchanges with D but not with F.

Since the emergence of power use is expected to occur over time, only data from the last trial block were used to test hypothesis 2b. As predicted by this hypothesis, E's profits in exchanges with D were significantly greater than 12 ($\bar{X} = 15.26$, $t = 2.75$, $df = 19$, $P < .01$, one-tailed test). Similarly, E was also able to obtain significantly greater than 12 points in exchanges with F ($\bar{X} = 16.18$, $t = 6.46$, $df = 19$, $P < .01$, one-tailed test). Thus, hypothesis 2b received clear support.

Hypothesis 5 implies that the profits received by E should be greater under conditions of high incentive than low incentive in exchanges with

TABLE 1

EXPERIMENTAL RESULTS:
MEAN PROFIT OF PERSON E PER EXCHANGE WITH D AND WITH F
in NETWORK 1c BY EXCHANGE INCENTIVE AND TRIAL BLOCK

EXCHANGE INCENTIVE AND EXCHANGE PARTNER	TRIAL BLOCKS		
	1	2	3
Low:			
D	13.80 (4.13)	12.69 (4.26)	13.32 (4.25)
F	13.27 (3.10)	14.78* (2.77)	15.44** (2.96)
High:			
D	12.90 (3.71)	13.72 (4.40)	17.19** (5.26)
F	15.52** (2.38)	16.66** (2.10)	16.91** (2.46)
Combined:			
D	13.35 (3.95)	13.21 (4.36)	15.26** (5.16)
F	14.40** (2.99)	15.72** (2.63)	16.18** (2.82)

NOTE.—The profit obtained by D and F in negotiations with E can be obtained by subtracting the values in this table (E's profit) from 24. Standard deviations are in parentheses.
* Significantly greater than 12 ($P < .05$).
** Significantly greater than 12 ($P < .01$).

[17] This result may be due to a "ceiling" effect. The exchange rate seems to stabilize in the high-incentive condition at about 16 to eight by trial block 2, approaching but not reaching its theoretical maximum. This same rate of exchange was also approached in the low-incentive condition but stabilized later (i.e., in the third trial block).

both D and F. To test this aspect of the hypothesis, E's profits in the low-incentive condition were compared with E's profits in the high-incentive condition. For E's exchanges with D, this difference was significant ($\bar{X}$ = 3.87, t = 1.72, df = 18, $P < .05$, one-tailed test). However, for E's exchanges with F, the difference was not significant ($\bar{X}$ = 1.47, t = 1.15, df = 18, N.S.). Once again, hypothesis 5 was supported for exchanges with D but not with F (see n. 17).

Differential Emergence of Power Use over Time

Hypothesis 3 predicted that the power use of E over F (the most peripheral position) would be displayed first, followed by the emergence of power use by E over D (the most central position), because occupants of position F are more dependent initially than the occupants of position D. To test this hypothesis, a difference score was computed by subtracting the points obtained by E in exchanges with F from the points obtained by E in exchanges with D for the first trial block. A t-test for correlated means indicated that this difference was not significant ($\bar{X}$ = 1.04, t = 1.17, df = 19, N.S., one-tailed test). However, by the second trial block, when power use had begun to emerge, the difference was significant ($\bar{X}$ = 2.51, t = 3.19, df = 19, $P < .01$, one-tailed test). Thus hypothesis 3 was supported. (This hypothesis makes no claim about *when* power use will emerge; it claims only that it will emerge earlier in exchanges with F than in those with D.)

Hypothesis 5 implies that the differential emergence of power use of E over F and D predicted in hypothesis 3 will occur earlier under conditions of high exchange incentive than under conditions of low incentive. To test this hypothesis, a difference of differences score was calculated (i.e. $[\text{Profit}_{EF} - \text{Profit}_{ED}]_H - [\text{Profit}_{EF} - \text{Profit}_{ED}]_L$, where the first difference is obtained from the high-incentive condition and the latter from the low-incentive condition). For the first trial block this score was significant ($\bar{X}$ = 3.14, t = 1.88, df = 18, $P < .05$, one-tailed test); it was not significant for the second trial block ($\bar{X}$ = 0.84, t = 0.52, df = 18, N.S.). This finding suggests that the differential in power use by E over F relative to E's use of power over D emerged earlier under conditions of high incentive than under low-incentive conditions as implied by hypothesis 5.

Equal Powerlessness of the Central and Peripheral Positions

Hypothesis 4 predicted that, in the final phases of the exchange process, occupants of the most powerful position, E, would be exercising an equal amount of power use over both D and F. This implies that the profits E

obtains from D will not be significantly different from those E obtains from F in the final trial block. The hypothesis assumes, however, that the exchange process has stabilized. As a check on this assumption, E's profits for the first and second halves of the final trial block were compared. If the exchange process has stabilized by the last trial block, there should be no significant difference between E's profits in the first and second halves of this final exchange phase. For the high-incentive condition, this was true ($F = 0.01$, $df = 1,16$, N.S.). However, the difference was close to significant in the condition of low exchange incentive ($F = 3.80$, $df = 1,16$, $P < .07$). This suggests that under conditions of low incentive, the exchange process may not yet have stabilized; hence we shall test hypothesis 4 only for the high-incentive condition. A t-test for correlated means revealed no significant difference between the amount of profit E received in exchanges with D and that received in exchanges with F in the high-incentive condition ($\bar{X} = 0.28$, $t = 0.19$, $df = 18$, N.S.); the test confirms hypothesis 4.

Hypothesis 5 implies in addition that the equality in the level of E's use of power over D vs. F will emerge more quickly under conditions of high than of low incentive. To test this implication, E's profit from exchanges with D versus F under high incentive was compared with E's exchange profits from D versus F under low incentive. This "difference of differences" approached significance, indicating that the predicted equality of power use by E over D versus F (i.e., hypothesis 4) tended to be achieved earlier under the condition of high exchange incentive ($t = 1.38$, $df = 18$, $P < .10$, one-tailed) than under that of low exchange incentive.

These results taken together provide strong support for the basic predictions derived from power-dependence reasoning concerning the distribution of power over time in negatively connected exchange networks. All of the primary hypotheses (i.e., hypotheses 2–4) were supported. Hypothesis 5, concerning the effects of differential levels of exchange incentive, was partially supported. In general, the effects predicted in hypotheses 2–4 were stronger under conditions of high exchange incentive than under those of low exchange incentive. Even when the predicted differences did not obtain (e.g., for hypothesis 2), the results suggest that the effect of low exchange incentive was to delay E's use of power over D, whereas F's dependence was so great that the exchange incentive made no difference in the emergence of E's use of power over F (see table 1). The emergence of E's power over D is, in theory, predicted to occur subsequent to the emergence of power over F (see results for hypothesis 3). Thus, E's power differential over D may not have had sufficient time to emerge in the low-incentive condition before the end of the experiment. The findings concerning the lack of stabilization (see hypothesis 4 results) of

the exchange process during the final trial block under low incentive further support this contention.

COMPUTER SIMULATION OF MORE COMPLEX EXCHANGE NETWORKS

Having demonstrated empirically the predictive power of power-dependence theory within network $1c$, we turn now to predicted differences in power use among the networks in figure $1c$–f. Hypotheses 6, 7, and 8 present these predictions. Because laboratory experiments with such large networks are too costly to conduct, we developed a computer simulation to "test" these hypotheses. Because hypotheses 1–4 apply to all of the more complex networks as well as to network $1c$, the generality of those predictions was also explored in the simulation.

A program (labeled SIMNET) was written to simulate the negotiation of exchange among actors in any negatively connected N-actor network.[18] The program has the capacity to vary the size and "shape" of the network, the number of trials, the amount of profit available for various exchange relations, the number of offers and counteroffers permitted within a given trial, and the "toughness" of the simulated actors (i.e., their tendency to drive relatively hard bargains). For consistency with the assumptions underlying power-dependence theory, the simulated actors were programmed to act "rationally," that is, to attempt to maximize their profits through the exchange process. As in the theory presented above, the "power" of these simulated "actors" can derive only from their location in the network, which links them to identically programmed other "actors."

"Rationality" in this bargaining program means specifically that each actor: (1) accepts the better of any two offers, (2) raises "its" demand the next time if its offer has been accepted, and (3) lowers its demand when an offer goes unaccepted. When an actor receives an offer which is greater than the one it is currently seeking, the actor increases its demand the next time; it decreases its demand when the incoming offers are lower than its own past demands. The initial offers (or demands) sent out by the simulated "actors" at the beginning of the first trial were randomly assigned within the range of 1–23. Initial offers of 0 and 24 were not allowed, nor were negative demands. The initial offer was sent to all the exchange partners for any given actor. At the beginning of subsequent trials, the initial demand of each actor was increased if an exchange had been completed on the previous trial or decreased if the actor had failed to complete an exchange on that trial.

---

[18] The simulation program was written in FORTRAN on a PRIME 300 mini-computer by one of the authors, Toshio Yamagishi. Details concerning the program can be obtained by writing this author.

To test our predictions, we simulated the 5-, 7-, 10-, and 13-actor networks in figure 1*c–f*. The number of transaction periods, or trials, was twice that used in the experiment, or 54 trials, to allow us to examine any trends which might be produced over a longer period of time. For each network structure, 50 replications were conducted, so that $N = 50$. We present the results in table 2, which shows the average points obtained by E_i from F_j and D in each network structure across six trial blocks of nine trials each.

Simulation Results

Simulation results can be examined in two ways: (*a*) for evidence that the simulation program is realistic, and, if there is such evidence, (*b*) as "data" to support or contradict our hypotheses.

a) Simulation of the five-person network in figure 1*c* allows a direct comparison with the results obtained from human subjects. The simu-

TABLE 2

SIMULATION RESULTS:

MEAN PROFIT OF THE POWERFUL (E) PER "EXCHANGE" WITH D AND WITH F IN FOUR REPLICATIONS VARYING NETWORK SIZE

FIGURE PART, SIZE OF NETWORK, AND E's EXCHANGE PARTNER	TRIAL BLOCKS					
	1	2	3	4	5	6
1*c*, 5 actor:						
D	12.79a	15.36	17.15	18.60	19.31	19.55
	(2.57)	(2.76)	(2.98)	(2.11)	(1.67)	(1.44)
F.	14.71	16.33	17.83	19.08	19.86	19.91a
	(2.64)	(2.76)	(2.56)	(1.86)	(1.42)	(1.26)
1*d*, 7 actor:						
D	10.56b	13.03	15.00	16.54	17.64	18.47
	(3.02)	(3.07)	(2.81)	(2.32)	(1.68)	(1.43)
F.	14.33	15.31	16.64	17.79	18.66	19.06
	(2.95)	(2.34)	(2.40)	(1.93)	(1.38)	(.97)
1*e*, 10 actor:						
D	13.99	17.69	19.65	20.06	20.11	20.11
	(2.89)	(2.37)	(1.07)	(.58)	(.55)	(.54)
F.	16.35	18.68	19.86	20.11	20.22	20.01
	(2.16)	(1.35)	(.59)	(.42)	(.43)	(.39)
1*f*, 13 actor:						
D	14.50a	19.56c	20.42	20.63	20.50	20.43
	(3.18)	(2.10)	(.86)	(.66)	(.64)	(.55)
F.	17.18	20.06	20.67	20.87	20.69	20.58
	(1.69)	(.74)	(.57)	(.59)	(.52)	(.49)

NOTE.—These values represent the average profit E obtained in "exchanges" with D and F, with 24 units of profit available for each "exchange"; therefore D's and F's average profit equals 24 − E's profit in each case. Each trial block contained nine trials. Cell values are based on the simulation of 50 groups; in an occasional group, however, E did not complete an "exchange" in a given trial block. Cell means labeled "a" are based on 49 groups, that labeled "b" has 47 groups, that labeled "c" has 42 groups per cell; all others have 50 groups per cell. Standard deviations are in parentheses.

lation results in table 2 reproduced the entire pattern of experimental results in the high-incentive condition. (The low-incentive condition was not simulated.) Hypotheses 2, 3, and 4 are supported both by real and by simulated subjects. These parallel findings can be used to infer both the "rationality" of our real subjects and the "realism" of SIMNET.

b) An examination of table 2 shows that all of the relevant hypotheses advanced through the application of power-dependence theory were supported by the simulation results. Specifically, the power of E over D emerged more slowly in network 1*d* than in 1*c* where D had fewer alternatives, as predicted by hypothesis 6. However, when the number of alternatives for D is constant, the rate at which equilibrium occurs depends on the number of alternatives for E, as indicated by hypothesis 8 and supported by the data in table 2 for the relevant networks 1*d–f*. Similarly, as predicted by hypothesis 7, the power of E over F emerged more quickly as the number of E's alternatives increased (networks 1*d–f*). Thus, while D and F$_j$ are shown to be equally powerless in the long run, the rate at which this equilibrium condition was achieved differed systematically as specified by hypotheses 6–8.

DISCUSSION: POWER, DEPENDENCE, AND CENTRALITY

The findings obtained both from human subjects in a simple network (fig. 1*c*) and from the simulation of more complex networks demonstrate clear support for the predictions based on power-dependence theory. In contrast, two of the best conventional measures of point centrality (closeness and betweenness) fail to generate sound predictions concerning the distribution of power in negatively connected exchange networks. As a result of these findings, we arrive at two major conclusions. First, if we are to retain the intuitively appealing notion that power is a function of centrality, we must either develop a more general conception of centrality or apply current measures of point centrality only in certain types of networks.[19] For reasons of theoretical parsimony and generality, the former strategy is preferred. Second, while we have shown that power-dependence theory provides a very good basis for predicting the distribution of power in these networks, the theory was originally formulated at a very micro level ill-suited to the analysis of complex network structures. Therefore, power-dependence theory needs to be raised, if possible, to a more macroscopic level of analysis.

[19] One solution is to specify theoretically the conditions under which different measures of centrality apply. Freeman (1979) has begun this task. However, the logical conclusion to such efforts might well be the increased proliferation of centrality measures. For the sake of parsimony, it would be preferable to develop more general conceptions and measures of centrality.

Power-dependence theory (Emerson 1962, 1972) examines the power of one actor over another on the basis of the dependence of the latter. It is therefore fundamentally dyadic. What is needed is the determination of power at a position within a structure, on the basis of the "dependence" of the entire structure on that position. We suggest (a) that a measure of such system-wide dependence on a given position in the network will turn out to be a measure of the "centrality" of that location and (b) that power at this location can be interpreted easily in power-dependence terms. Furthermore, we suggest that such a dependency-based concept of centrality may be general, applying across all types of exchange networks, whether negatively connected, positively connected, or mixed. In this section of the discussion we can only point out the basic features of this approach, leaving a complete formulation for later work.

Dependence and Network Vulnerability

We start with this basic question: to what extent does the flow of valued resources (information, economic goods, political patronage, etc.) within an N-actor network depend on facilitating exchange behavior by the occupants of a given position in that network? Stated differently, to what extent will reduced participation or exchange activity at a given location have detrimental consequences for exchange throughout the network?

Our first step toward a theoretical solution to such questions was prompted by the graph-theoretic concept of "vulnerability"—the vulnerability of the network (or graph) to the removal of a given point or line (Harary et al. 1965). By "removal" we mean substantively any form of withdrawal from exchange activity. To remove a point (P$_i$) from a graph (G) is to obtain a subgraph called a "residual graph" (RG$_i$). Compared with the parent graph (G), the structure of the residual graph (RG$_i$) might be "weakened" or impaired in terms of resource flow, in which case the parent graph G is said to be "point vulnerable" at P$_i$. We suggest here that the contribution of P$_i$ to the network G and the "dependence" of exchange in that network on P$_i$ can be studied by comparisons between G and RG$_i$. Since we have defined a position as a set of points whose residual graphs are isomorphic, there need only be as many G and RG$_i$ comparisons as there are positions in the network structure (i.e., residual graphs for points occupying the same position in a network are identical). These comparisons will provide a measure of the "dependence" of the network as a whole on each position therein.

While a large number of more or less complex and refined measures can be derived from comparison of a graph with its residual graphs, one simple measure using the network in figure 1c (reproduced below for convenience) will serve to illustrate.

299

With this network taken as graph G, the three residual graphs shown in figure 2 are formed by the removal of points from positions D, E, and F, respectively. In our experiment, 24 resource units were exchangeable along solid lines and eight units along broken lines. From that information we can calculate what we refer to as the Reduction in Maximum Flow (RMF) in the total network which would result if a given point were removed. The results are shown in table 3. By this measure our laboratory network appears to be vulnerable only at position E, the position shown to be most powerful both in the actual experiment and in the simulation findings.

Vulnerability in a negatively connected network locates the points of minimum dependence, equivalent to maximum network-wide power. Even though in these networks there are no "indirect" paths of resource flow
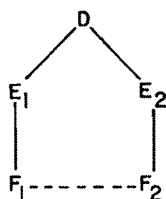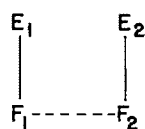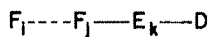

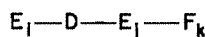FIG. 1c.—Five-person network (three positions)

RG$_d$

(Point D Removed)

RG$_e$

(Point E$_i$ Removed)

RG$_f$

(Point F$_i$ Removed)

FIG. 2.—Residual graphs formed by the removal of points from positions D, E, and F respectively in network 1c (graph G).

TABLE 3

REDUCED MAXIMUM FLOW AS A MEASURE OF NETWORK VULNERABILITY
AT EACH POSITION IN THE EXPERIMENTAL NETWORK 1c

Graph	G	RG$_d$	RG$_e$	RG$_f$
Maximum flow rate*.............	48	48	32	48
Reduced maximum flow (RMF) ..	...	0	16	0

* Maximum resources exchangeable in the network per transaction period.

as there would be in a positively connected network, the volume of re-source flow within exchange relations is dictated both by accessibility to exchange partners and by the availability of resources from those partners. Thus positions are relatively "powerless" in a network (e.g., D in network 1c) to the extent that they have few exchange opportunities (i.e., few alternative sources of valued resources) and have direct connections only to actors who have highly reliable alternative sources of supply. In any network of exchange, availability of resources from exchange partners is critically determined by the nature of their connections to other sources of benefit. Thus while there are no indirect paths of resource flow (e.g., indirect exchange) in negatively connected networks, there are indirect effects of the structure of the alternative exchange opportunities, and these effects have repercussions throughout the network. These are struc-tural implications of the nature of the exchange connections. It is im-portant to note that the overall structure determines the distribution of power in negatively connected networks which do not involve resource flows across the entire network. The distribution of benefit, however, is dictated by the distribution of power even though the actors have no awareness of the total potential benefit to be obtained in the system through exchange activity.[20]

We have developed only a first approximation of a general measure of network-wide dependence on a given actor (or point).[21] This discussion is meant only to illustrate the theoretical potential of a "vulnerability" approach to the problem of raising power-dependence theory from a dyadic to a more macrostructural level of analysis. At a minimum this conception achieves identification of the centrally located positions in networks which are negatively connected.

System-Level Dependence and Centrality

This conception of vulnerability can also be seen as a useful approach to the measurement of point centrality in an exchange network that is sen-sitive to the nature of the connections that join dyads into networks. If the networks depicted in figure 1c and e, for example, were positively connected at E_1 (i.e., if E_1-F_1 exchange were contingent on E_1-D exchange), the removal of D would greatly reduce the network capacity for resource

[20] Since the actors do not have knowledge of the total potential gain possible through exchange activity, the exercise of power by any actor in the structure is a function of position in the network and not of any direct knowledge or awareness of his potential to thwart the "efficient" (or maximizing) distribution of resources across the network.

[21] The RMF measure requires modification if it is to be applied to digraphs (see Yamagishi 1981).

flow[22] (i.e., D is a point of vulnerability in this network). For E_i to continue to receive resources of value from F_i in the network, E_i would become dependent on D to maintain the flow of these resources to position E. Thus resource flow in the network as a whole would be highly dependent on D, returning power to the "center" in this positively connected network.

A "betweenness" measure of point centrality in positively connected networks is consistent with this vulnerability notion (as a measure of network-wide dependence on a point). As Freeman (1979) notes, what is at issue theoretically in communication networks is the potential for control through "withholding or distorting information." By extension to positively connected exchange networks, this concept would refer to the potential control over the rate of resource flow (i.e., through the potential to withhold resources or the failure to transmit them to other exchange partners in the network). While space will not allow a complete analysis of positively connected networks here, it does appear that the general notion of "vulnerability" can be adapted to positively connected as well as negatively connected networks (and, by implication, to mixed networks). What is interesting is that dependence and centrality clearly converge in this theoretical approach, as the choice of wording by Marsden and Laumann (1977, p. 224) would suggest. Further theoretical development and an experiment on positively connected networks are now in progress.

CONCLUSIONS

In this article we have dealt primarily with negatively connected networks, in an extension of previous research on the distribution of power in such networks. Our empirical research suggests that a very interesting structural principle applies to exchange networks of this type. It can be referred to as a "decentralization" principle: such networks tend to form into systems organized around multiple foci of power at the points E_i in figure 1. Those points can be viewed as "regional centers" of power, like petty kingdoms in an encompassing empire (see Emerson 1982). Those points are defined as "central" if centrality is measured in terms of network-wide vulnerability at point P_i (e.g., by a measure like the RMF measure developed here). But such points cannot be considered "central" in any sense of the term by any of the existing measures of point centrality without falling into circular reasoning concerning the relation of power to centrality.

[22] In a network of direct exchange relations like the one in fig. 1a, the removal of the central point (e.g., A) completely halts profitable exchange activity, since the actors on the periphery cannot engage in a profitable exchange of resources.

Thus our research has identified a major weakness in existing point-centrality measures: they are not applicable to negatively connected exchange networks. To fill this gap, we developed a measure based on the concept of vulnerability, network-wide dependence on a particular point. More important, we have suggested that this general notion may be developed to apply to positively connected networks as well, since in networks of this type vulnerability seems to correspond to the underlying theoretical meaning of betweenness–based measures of point centrality (see Freeman 1979). Thus, it is reasonable to expect that in positively connected networks, "centralization" (i.e., a power shift to the center) is more likely to occur than decentralization because of the network-wide dependence on point D (if the connections are defined as positive instead of negative in fig. 1c–f). Position D in such networks serves as the only resource link among the various peripheral subsystems of exchange activity (e.g., E_i-F_i).

These notions are being developed further in order to specify theoretically the implications of the different types of exchange connections. It is clear that the integration of structural network principles with exchange network theory provides useful insights into the dynamics of power in networks of connected exchange relations. This type of theoretical activity will not only extend exchange theory but also provide one potential theoretical basis for network theory (see Cook 1982). Finally, this theoretical formulation offers an explicit procedure for linking actors' exchange behavior to network properties (Foster 1979) and suggests mechanisms which may yield "possible transformations" of these networks as a result of power dynamics or changes in the nature of the exchange connections.

REFERENCES

Bavelas, A. 1948. "A Mathematical Model for Group Structures." *Human Organization* 7:16–30.
———. 1950. "Communication Patterns in Task Oriented Groups." *Journal of the Acoustical Society of America* 22:271–82.
Berge, C. 1962. *The Theory of Graphs*. New York: Wiley.
Blalock, H. M., and P. H. Wilken. 1979. *Intergroup Processes: A Micro-Macro Perspective*. New York: Free Press.
Burt, R. S. 1976. "Positions in Networks." *Social Forces* 55:93–122.
Busacker, R., and T. Saatz. 1965. *Finite Graphs and Networks: An Introduction with Applications*. New York: McGraw-Hill.
Cook, K. S. 1978. "Exchange, Power and Equity." Paper presented at a Didactic Seminar at the 74th annual meeting of the American Sociological Association, Boston.
———. 1982. "Network structures from an exchange perspective." Pp. 177–99 in *Social Structure and Network Analysis*, edited by P. Marsden and N. Lin. Beverly Hills, Calif.: Sage.
Cook, K. S., and R. M. Emerson. 1977. "Design of a Computerized Laboratory for the Experimental Analysis of Social Interaction." Technical Report. Seattle: University of Washington.

————. 1978. "Power, Equity and Commitment in Exchange Networks." *American Sociological Review* 43:712–39.

Emerson, R. M. 1962. "Power-Dependence Relations." *American Sociological Review* 27:31–40.

————. 1972. "Exchange Theory, Part II: Exchange Relations and Networks." Pp. 58–87 in *Sociological Theories in Progress,* edited by J. Berger, M. Zelditch, and B. Anderson. Vol. 2. Boston: Houghton Mifflin.

————. 1981. "Social Exchange Theory." Pp. 30–65 in *Social Psychology: Sociological Perspectives,* edited by M. Rosenberg and R. Turner. New York: Academic Press.

————. 1982. "Charismatic Kingship in Baltistan." Mimeographed. Seattle: Department of Sociology, University of Washington.

Foster, Brian L. 1979. "Formal Network Studies and the Anthropological Perspective." *Social Networks* 1:241–56.

Freeman, L. C. 1979. "Centrality in Social Networks: Conceptual Clarification." *Social Networks* 1:215–39.

Granovetter, M. 1979. "The Theory-Gap in Social Network Analysis." Pp.501–18 in *Perspectives on Social Network Research,* edited by P. Holland and S. Leinhardt. New York: Academic Press.

Harary, F., R. Norman, and D. Cartwright. 1965. *Structural Models.* New York: Wiley.

Homans, G. C. 1961. *Social Behavior: Its Elementary Forms.* New York: Harcourt, Brace.

————. 1974. *Social Behavior: Its Elementary Forms.* Rev. ed. New York: Harcourt, Brace.

Hopkins, T. 1964. *The Exercise of Influence in Small Groups.* Totowa, N.J.: Bedminster.

Laumann, E. O., and F. U. Pappi. 1973. "New Directions in the Study of Community Elites." *American Sociological Review* 38:212–30.

————. 1976. *Networks of Collective Action.* New York: Academic Press.

Leavitt, H. J. 1951. "Some Effects of Certain Communication Patterns on Group Performance." *Journal of Abnormal and Social Psychology* 46:38–50.

Mackenzie, K. D. 1966. "Structural Centrality in Communication Networks." *Psychometrika* 31:17–25.

Malinowski, B. 1922. *Argonauts of the Western Pacific.* New York: Dutton.

Marsden, P. V. 1981a. Introducing Influence Processes into a System of Collective Decisions." *American Journal of Sociology* 86:1203–35.

————. 1981b. "Restricted Access in Networks and Models of Power." Paper presented at the Conference on Structural Sociology, Albany, N.Y., April.

————. 1982. "Brokerage Behavior in Restricted Exchange Networks." Pp. 201–18 in *Social Structure and Network Analysis,* edited by P. Marsden and N. Lin. Beverly Hills, Calif.: Sage.

Marsden, P., and E. O. Laumann. 1977. "Collective Action in a Community Elite." Pp. 199–250 in *Power, Paradigms and Community Research,* edited by R. J. Liebert and A. W. Imershein. Beverly Hills, Calif.: Sage.

Miller, J. 1980. "Access to Interorganizational Networks as a Professional Resource." *American Sociological Review* 45:479–96.

Mizruchi, M. S., and D. Bunting. 1981. "Influence in Corporate Networks: An Examination of Four Measures."*Administrative Science Quarterly* 26:475–89.

Moreno, J. L. 1934. *Who Shall Survive?* Washington, D.C.: Nervous and Mental Disease Publishing Co.

————. 1943. "Sociometry and the Cultural Order." *Sociometry* 6:299–344.

Nieminen, J. 1974. "On Centrality in a Graph." *Scandinavian Journal of Psychology* 15:322–36.

Rogers, D. L. 1974. "Sociometric Analysis of Interorganizational Relations: Applications of Theory and Measurement." *Rural Sociology* 39:487–503.

Sabidussi, G. 1966. "The Centrality Index of a Graph." *Psychometrika* 31:581–603.

Sarason, S., C. Carrol, K. Maton, S. Choen, and E. Lorentz. 1978. *Resources, Community and Exchange Networks.* San Francisco: Jossey-Bass.

Shaw, M. E. 1954. "Group Structure and the Behavior of Individuals in Small Groups." *Journal of Psychology* 38:139–49.

————. 1964. "Communication Networks." Pp. 111–47 in *Advances in Experimental Social Psychology.* Vol. 6. Edited by L. Berkowitz. New York: Academic Press.

White, H. D., S. A. Boorman, and R. L. Breiger. 1976. "Social Structure from Multiple Networks. I. Blockmodels of Roles and Positions." *American Journal of Sociology* 81:730–80.

White, H. D., and R. L. Breiger. 1975. "Patterns across Networks." *Transaction* 12:68–73.

Yamagishi, T. 1981. "Two Approaches in Defining Positions in Networks." Mimeographed. Seattle: Department of Sociology, University of Washington.

Multiple Discoveries and the Central Message[1]

Don Patinkin
Hebrew University of Jerusalem and *Maurice Falk Institute for Economic Research in Israel*

Discussions of independent multiple discoveries in science have paid inadequate attention to two essential points: precise definition of the discovery and the extent to which it was part of the scientist's central message. When account is taken of these points, some cases of alleged multiple discovery turn out to be singletons. This contention is illustrated by means of an examination of the alleged multiple discovery of the theory presented by J. M. Keynes in his *General Theory of Employment Interest and Money*. The analysis indicates that the familiar claims of the Stockholm school and of Michał Kalecki in this regard should be rejected. The paper ends with a discussion of the importance from the viewpoint of the sociology of science that a discovery be part of a scientist's central message.

In his well-known paper on the phenomenon of independent multiple scientific discoveries, Robert Merton (1961, p. 356) contended that this phenomenon constitutes the "dominant pattern" in science: that indeed "all scientific discoveries are in principle multiples." On the basis of a detailed case study, I shall contend in this paper that in at least some instances such multiplicity disappears once one (1) provides a precise definition of the discovery in question and (2) examines the extent to which the alleged co-discoverers "really meant it"—the extent to which they incorporated the discovery in what I shall call their "central message." Indeed, my central message is that, from the viewpoint of science as an institution for the discovery of knowledge, a scientist should not be regarded as having made a discovery unless it is part of his central message.

The question of multiple discovery is related to the broader question of the relative importance in scientific discovery of the individual scientist, on the one hand, and of historical processes, on the other. Clearly, both play a role. As Barber (1952, p. 265) aptly put it, "There emerge multiple discoveries by men whose activity is guided in part by the existing scientific heritage and in part by their creative imaginations." And Merton (1961, pp. 366–69) has himself emphasized that his hypothesis does not imply that all discoveries are the inevitable product of historical processes. At the same time, it is reasonable to say that the more scientific development is characterized by multiple discovery, the less important is the role of the individual scientist.

The case I shall examine is the alleged independent multiple discovery of the theory which John Maynard Keynes presented in his celebrated *General Theory of Employment Interest and Money* (1936). (For convenience, I shall refer to this theory as the General Theory—not italicized.) A short time after the appearance of Keynes's book, Bertil Ohlin ([1937] 1944, p. 87) claimed that there were "surprising similarities" between the analysis in this book and that which had been developed in the Swedish writings of what he called the "Stockholm school," under which rubric he included Erik Lindahl and Gunnar Myrdal as well as himself. Similarly, in a review article on Keynes's book, the Polish economist Michał Kalecki (1936) claimed that he had anticipated its main arguments in a 1933 monograph in Polish. Ohlin's claim was presented in the *Economic Journal,* then the leading journal of the economics profession, and gained immediate attention—so much so that the claim of the Stockholm school became a "perennial of doctrinal history" (in Gustafsson's [1973] apt phrase). In contrast, Kalecki's claim was published in *Ekonomista*—the professional journal of Poland's economists, published, of course, in their own language—and so received no attention outside that country. Fifteen years later, however, the claim for Kalecki was brought to the attention of the profession as a whole by Lawrence Klein (1951) and Joan Robinson (1952)—and has received increasing support ever since.

As a preliminary, let me note that, whether we consider the progress of science to be determined by external social needs or whether we consider it to proceed by a dynamics of its own (Laudan 1977, chap. 7), there are good a priori reasons why we might expect the General Theory to have been discovered simultaneously by several economists, even by economists in different countries. From the viewpoint of social needs, the pernicious problem of persistent depression and unemployment which formed the background of the General Theory was common to the entire Western world in the early 1930s; from the viewpoint of internal dynamics, the state of economics as a science in several of these countries (e.g., Britain, the United States, Sweden) did not basically differ at that time.

The question which concerns me in this paper is whether the consequent a priori expectation of a multiple discovery of the General Theory is indeed valid.

In accordance with my opening remarks, let me first describe Keynes's major contribution—his central message: we cannot meaningfully discuss the possible multiple discovery of the General Theory unless we first specify what this contribution was. Clearly, the broader this specification, the greater the likelihood of finding the contribution anticipated. Thus, if we define the contribution as lying in the notion of aggregate demand coupled with the contention that aggregate demand can fall short of aggregate supply and thus generate unemployment, then—as Keynes himself noted in the *General Theory* (pp. 362–71)—the notion is to be found (albeit not fully developed) in the writings of Malthus in his famous debate with Ricardo more than a century before on the latter's contention that there could not exist a "general glut on the market" (i.e., Say's law). Similarly, if we identify Keynes's contribution as lying in his emphasis on the crucial role of fluctuations in investment in generating business cycles in a capitalist economy, then this theme is to be found at least as far back as the business-cycle theories of Tugan-Baranowsky (who in turn was influenced by Marx) and Spiethoff at the turn of the century (Hansen 1951, pp. 277–300). And if we see Keynes's contribution as lying in his advocacy of public works as a means of combating unemployment, then such policies had been advocated in Britain, Sweden, and other countries before World War I (Winch 1969, pp. 53–57; Ohlin 1981, p. 191). More to the point, by the end of the 1920s in Britain, as the unemployment generated by the return to the gold standard in 1925 at prewar parity dragged on, most British economists (and Keynes in particular) were already advocating public works as a remedy, albeit without much success (Keynes and Henderson 1929; Hutchison 1953, pp. 409–23, 1968, pp. 277–79; Winch 1969, pp. 104–46, 145–97, 343). Furthermore, Keynes and Henderson (1929), loosely, and subsequently Richard Kahn (1931), rigorously, had shown that by virtue of the increased output and hence income which it generated, such investment in public works would also generate the increase in savings necessary to finance the initial increase in investment—thus refuting the "Treasury view" that the increased public investment would simply reduce the amount of savings available for private investment, so that no net increase in investment expenditures and hence employment would take place.

What then was Keynes's central message? Like many other students of his writings, I consider it to be the theory of effective demand as an explanation of a state of "unemployment equilibrium" in a capitalist economy. In figure 1, this theory is presented diagrammatically under simplifying assumptions. Its essence is not only that the intersection of the

aggregate-demand curve $E = F(Y)$ with the 45° line representing aggregate supply determines equilibrium real output Y_o at a level that may be below that of full employment Y_F; not only that disequilibrium between aggregate demand and supply causes a change in output and not in the price level (on which more in a moment); but also (and this is the distinctively novel feature) that the change in output itself acts as an equilibrating force. That is, if the economy is at (say) the level of output Y_1, characterized by an excess of supply (Y_1B) over demand (Y_1A), then the resulting decline in output, and hence income, will (because of Keynes's crucial assumption of a marginal propensity to spend that is less than unity) reduce supply more than demand, thus reduce the initial excess supply, and thus eventually bring the economy to equilibrium at Y_o. Or, in terms of the savings = investment equilibrium condition (which can be shown to be equivalent to the foregoing demand = supply condition), the decline in income will decrease savings and thus eventually eliminate the excess of savings over investment that exists at Y_1. This was Keynes's explanation of the seemingly endless state of mass unemployment from which the Western world was then suffering.

Needless to say, the actual exposition of the *General Theory* is far more detailed and complex. In it Keynes enlarges on the foregoing analysis of demand and supply in the commodity market and integrates it with a corresponding analysis of the money market. And the elaboration, refinement, and empirical testing of this new analytical framework consti-
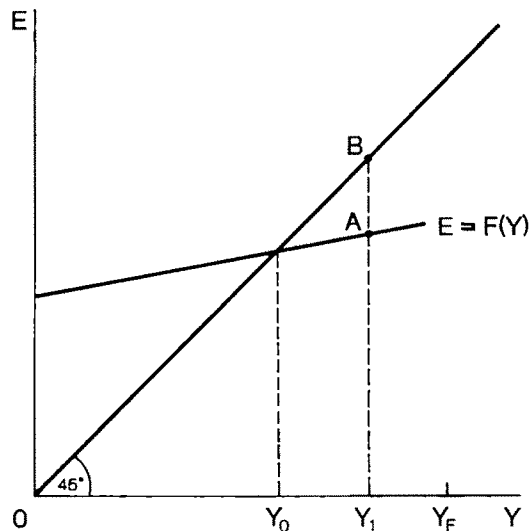

FIG. 1

tuted the research program of macroeconomics (to use Lakatos's [1970, pp. 132 ff.; 1978] term) for decades afterward. This was the nature of the "Keynesian revolution."

The foregoing diagrammatic analysis has long been a standard feature of introductory textbooks in economics. But that it was far from obvious in the early 1930s is best illustrated by the fact that just a few years before the *General Theory,* in his *Treatise on Money* (1930, 1:159–60), Keynes himself had argued that a decline in output generated by an excess of savings over investment would not automatically bring the economy to a new equilibrium position but would continue until it was possibly brought to an end by an adventitious decrease in savings or increase in investment. In brief, there is no recognition in the *Treatise* of the fact that the decline in output (and hence income) itself generates an automatic decrease in savings that ultimately brings them into equilibrium with investment.

A related aspect of the *Treatise* is that, although, like the *General Theory,* it assigns a critical analytical role to the relation between savings and investment, any difference between these two quantities (generated by a departure of the rate of interest from its equilibrium value) is assumed to reflect itself in the first instance not in changes in output but in changes in the price levels of consumption and investment goods, respectively. This was the central message that Keynes of the *Treatise* conveyed by means of what he termed his "fundamental equations." Correspondingly, although the *Treatise* is definitely concerned with changes in output and employment, it treats them as derivative from the changes in the price levels.

What must now be emphasized is that this concentration on the analysis of the determination of the price level (or, equivalently, its reciprocal, the purchasing power of money) was not unique to the *Treatise* but was instead a general characteristic of monetary theory before the *General Theory.* This can be seen from the very titles of the major works of that period. Thus the full title of Knut Wicksell's work (on which Keynes of the *Treatise* explicitly drew) was *Interest and Prices: A Study of the Causes Regulating the Value of Money* ([1898] 1936). Similarly, the full title of Irving Fisher's classic work on the quantity theory was *The Purchasing Power of Money: Its Determination and Relation to Credit Interest and Crises* (1911), while that of A. C. Pigou on the Cambridge version of this theory was entitled "The Value of Money" ([1917] 1951). Similarly, the full name of Keynes's equations in the *Treatise* was "The Fundamental Equations for the Value of Money."

With this as background, let me proceed to a discussion of the Stockholm school. The basic point here is that the central message of this school—like that of its intellectual father, Knut Wicksell—had to do, not with the equilibrating role of changes in output, but with the related

movements of the rate of interest and the various price levels and their subsequent effect on investment. Thus, though Lindahl ([1930] 1939, p. 142) presented his analysis in terms of an aggregate demand function for consumption goods, he assumed that the difference between this demand and the corresponding supply exerted its influence on the price level of such goods. And even when Lindahl did discuss possible changes in output, he did not take account of the possible equilibrating effects of such changes.

Similarly, the main objective of Myrdal's *Monetary Equilibrium* ([1931, 1933] 1939) was to provide a detailed critique of Wicksell's contention that there existed a "natural rate of interest" which would at one and the same time equal the marginal productivity of capital, equate investment and savings, and stabilize the price level (pp. 37–38). Again, the central message of Ohlin's major theoretical work ([1933*b*] 1978) prior to the *General Theory* had to do with the interrelated movements of interest and prices. Indeed, in the concluding section of this work Ohlin lists its five major objectives: the first of these has to do with the definition of income, and the remaining four with the behavior of prices and the interest rate; none of them deals explicitly with the behavior of output. As in the case of Keynes of the *Treatise*, this does not mean that Myrdal and Ohlin were not concerned with changes in output and hence employment: in the worldwide depression of the early 1930s no economist could be anything but concerned. But their analysis of changes in output derived from that of prices. Thus Ohlin ([1933*b*] 1978) did take some account of the feedback effects of changes in output on aggregate demand, but only with respect to the subsequent effect on prices and hence the business cycle; there was no recognition of the equilibrating role of changes in output.

Kalecki's 1933 essay[2] is closer in spirit to the General Theory in being primarily concerned with quantities rather than prices. But its central message has to do, not with unemployment equilibrium, but with the generation of cycles of investment (and hence of output and employment) in a capitalist economy on the implicit assumption that, at every instant of time in the cycle, planned savings and investment have—in some unspecified manner—been equilibrated. Thus Kalecki assumes away just that problem (namely, the nature of the equilibrating mechanism) which is the central message of the *General Theory*. Correspondingly, Kalecki summarizes his analysis graphically, not with a diagram that depicts a state of unemployment equilibrium, but with one that depicts the smooth cyclical movements over time of investment and hence output and employment. Similarly, the empirical evidence which Kalecki adduces in his 1933 essay has to do with the investment cycle.

[2] The "most essential part" of that essay appears in English translation as chap. 1 of Kalecki (1966). (See Kalecki's foreword [ibid., p. 1] for this characterization of the chapter.)

On what, then, have the respective claims of the Stockholm school and of Kalecki been based? In part on a different specification of the central message of the *General Theory,* and in part on discussions that I do not regard as part of their respective central messages. Thus, with reference to the first of these differentiae, the claims of these economists (or the claims made for them) have sometimes been based on the undeniable facts that, like Keynes, they too were concerned with the basic problem of unemployment and that they found its cause in the inadequacy of aggregate demand, generated in turn by an inadequacy in investment demand. Again, the respective cases for Ohlin and Kalecki have sometimes been based on their having both shown that—under conditions of unemployment—an increase in investment will increase output and hence income and thus generate the savings necessary for financing the initial investment. For reasons indicated above, however, neither of these points can be regarded as the central message of the *General Theory.*

I come now to my second and major differentia. As Steiger (1978, pp. 429, 440) has noted, there is indeed a passage in Myrdal ([1931, 1933] 1939, pp. 164–66, 169) which refers to the equilibrating role of a change in output. But because the major conclusion of this passage is concerned with prices and because the passage is contradicted by another one in the book which ignores this role (ibid., pp. 106–7), I do not regard this passage as part of Myrdal's central message. Further support for this view is provided by the fact that Myrdal made no attempt in his subsequent writings to develop the ideas of this passage in a more complete and precise manner. Indeed, he did not even advert to them in the relevant contexts of his 1934 report for the Swedish government, *The Economic Effects of Fiscal Policy.*

A similar reservation holds with respect to Kalecki. It is true that Kalecki does in effect discuss the Keynesian equilibrating mechanism in a 1935 article published in a Polish semigovernmental weekly magazine devoted to economic commentary and reports, *Polska Gospodarcza* [Economic Poland].[3] But as I have already noted, there is little if any concern with this mechanism in Kalecki's professional writings during the pre–*General Theory* period. And, lest I be misunderstood, let me emphasize that my point here is not that this theme appears in Kalecki's nonprofessional writings, but that it appears *only* there. (In a similar way, I would not consider as part of an economist's central message to the profession

---

[3] This article was translated many years later under the title "The Mechanism of the Business Upswing" (Kalecki 1966, pp. 2, 26–33). The masthead of *Polska Gospodarcza* stated that it was published with the support of the Ministry of Industry and Trade as well as the Ministry of Finance and other government ministries. For clarification of the nature of this magazine, I am indebted to Alexander Erlich, who has likened it to the business section of the *New York Times.*

a theory which he presented in, say, a column in *Newsweek* and which had no counterpart in his scientific writings.) It is also significant that, when Kalecki put forth his claims for priority in his 1936 review article on the *General Theory*, he did so entirely on the basis of his 1933 monograph and made no mention of the 1935 article.

At this point let me note that there is no doubt that Keynes—with his worldwide reputation, and situated at Cambridge University, then *the* center of learning in economics—was in an ideal position to communicate his message to the profession as a whole. This fact has been emphasized by, for example, those who have attempted to explain Kalecki's failure to achieve recognition for the discovery of the General Theory by saying that in the early and mid-1930s he was an obscure Polish economist writing in his own language (Feiwel 1975, pp. 23–24, 29). But this cannot be the only reason. At the 1933 meetings of the Econometric Society in Leiden, Kalecki had the opportunity of presenting to the profession a paper (1935*b*) on what he presumably regarded as the essential points of his 1933 Polish essay, and this paper even drew the attention of two of the then-leading macroeconomists of the world, Jan Tinbergen (1935) and Ragnar Frisch (1935). But both these experts perceived the central message of Kalecki's paper to be (as he himself perceived it), not a theory of unemployment equilibrium, but a theory of the investment cycle. That was the theme that ran through their heads when they left the hall.

Keynes's advantageous position for making his message heard is even less convincing as an explanation of his success vis-à-vis the Stockholm school. After all, in 1929 Ohlin had no difficulty in gaining a worldwide hearing for his difference of opinion with Keynes on the transfer problem generated by post–World War I German reparations. Even more to the point is the significant and rapid impact on the profession as a whole of the central message of Ohlin's *Interregional and International Trade* (1933*a*) (namely, factor proportions as the explanation of comparative advantage and hence the pattern of trade). I do not underestimate the importance of the fact that these writings appeared in English under prestigious auspices: Ohlin's debate with Keynes in the pages of the *Economic Journal* and his book as a volume in the Harvard Economic Studies. Nevertheless, I contend that if the early 1930s writings of Myrdal and Ohlin in the *Ekonomisk Tidskrift* and in the reports for their government's Committee on Unemployment did not have the impact of Keynes's *General Theory*, it is not only because they were written in Swedish and hence not readily accessible to the profession as whole, but also because their central message differed substantively from that of the *General Theory*. Correspondingly, the passages in Myrdal's and Ohlin's writings which they or others have cited as constituting independent discovery of the General Theory failed to have an impact not only on the English-speaking part of the

profession but on the Swedish-speaking as well. Thus, though Tord Palander's ([1941] 1953, pp. 37–43; see also p. 15, n. 14) lengthy and detailed critique in the *Ekonomisk Tidskrift* of Myrdal's *Monetary Equilibrium* devotes considerable attention to the case in which prices (including factor prices) are constant and quantities vary, Palander does not mention Myrdal's reference in this context to the equilibrating role of changes in output. Nor does Brinley Thomas in his survey of Myrdal's writings (1936, pp. 86–102, passim). Similarly, the aforementioned passages in Myrdal's and Ohlin's 1933 writings did not stimulate a research program along these lines by Swedish economists, as evidenced by the lack of pertinent discussion in subsequent issues of the *Ekonomisk Tidskrift*.

This constitutes my main thesis: the phenomenon for which a desperate world in the early 1930s was seeking an explanation was the bewildering and seemingly endless depression that was creating untold misery and threatening political stability. This was the problem to whose solution Keynes—in contrast to the Stockholm school and Kalecki—devoted his central message: namely, that the state of unemployment equilibrium could be explained by his theory of effective demand, according to which the decline in output itself generates feedback effects which ultimately eliminate the excess of aggregate supply over aggregate demand. I might also note that the fact that this theoretical revolution occurred simultaneously with the revolution in national income statistics which made possible the quantification of Keynes's analytical categories further increased the impact of his *General Theory* on the profession (see Patinkin 1982, chap. 9).

I must now finally address the question of how the central message of a work can be identified and why I do not regard a scientist as having made a discovery unless it is part of this message. Admittedly, the literature that we have been examining (with the exception of Ohlin's 1933 article) does not generally provide a precise and unambiguous statement of its central message—of the major problem to which it addresses itself and the conclusions it reaches. And in the absence of such a statement I suggest that the central message is identified in much the same way that we identify the theme of a movement in a classical symphony: by its being announced early in the movement and by repetition afterward—either exactly or by different combinations of instruments or in different tonal areas. So the central message of a scientific work is announced by its presentation early in the work (and frequently in its title) and by repetition, either verbatim or modified in accordance with the circumstances.

Thus Lindahl ([1930] 1939) entitles the section of his work referred to above "The Rate of Interest and the Price Level"; Myrdal ([1931, 1933] 1939) announces that his study is an "immanent criticism" of Wicksell's

earlier work, which itself was concerned with the relation between interest and prices in an economy with full employment; Ohlin ([1933b] 1978) entitles the successive sections of his article "The Stability of the Price System," "The Morphology of the Price System," and "An Analysis of Some Price Movements"; Kalecki entitles his 1933 monograph "An Essay on the Theory of the Business Cycle"; and Keynes entitles his book *The General Theory of Employment Interest and Money* (in brief, everything but prices, which are "subsidiary" to his theory [p. 32]), presents in the introduction to the book a brief summary of the theory of employment "to be worked out in the course of the following chapters" (p. 72), and relegates his chapter of "Notes on the Trade Cycle" (a subject not even mentioned in the introduction) to the final part of the book, which is essentially an appendage to it.

And then there are repetitions. Indeed, the variety of relevant contexts for repeating the respective central messages of the works referred to in this paper is in some sense greater than that of a symphonic theme: for it includes repetitions not only in the work itself (prices in Ohlin's article, cycles in Kalecki's monograph, unemployment equilibrium in Keynes's *General Theory*) but in other works as well, some of them possibly by other writers. Thus in the preceding discussion I have identified the central message of a theoretical work by, inter alia, testing for its repetition by the writer in his related work of roughly the same period (Myrdal and Ohlin), in the related empirical studies that he may have carried out (Kalecki), in the part of his work that he chose for translation (Lindahl, Kalecki), and in the part of his work that he chose for presentation in leading scientific journals and at scientific meetings (Kalecki). And I have also tested for the repetition of the central message in contemporary comments on (Tinbergen and Frisch and Holme on Kalecki) or reviews of (Palander of Myrdal) the work in question, and in the writings of contemporaries who were exposed to the work (the messages of Myrdal and Ohlin as reflected in the writings of other Swedish economists in the *Ekonomisk Tidskrift*). On the basis of these criteria I have concluded that the respective central messages of the Stockholm school and of Kalecki were not that of the General Theory; that, in brief, the Stockholm message was concerned with prices and not output, and that Kalecki's was concerned with investment cycles and not unemployment equilibrium; and that, accordingly, the General Theory is not an instance of independent multiple discovery.

But why this insistence that a discovery be part of a writer's "central message"? If we are concerned with the question of multiple discovery, is it not enough that a scientist has made a discovery, no matter how incidental it may be to his main work?

To this question I have two answers, and both have to do with the role

of science as an institution for the discovery of knowledge. My first answer is a corollary of Merton's penetrating analysis (1957) of the importance of recognition of originality (as evidenced by chronological priority) in the reward system of science. Now, in Whitehead's oft quoted words ([1917] 1974, p. 127), "To come very near to a true theory, and to grasp its precise application, are two very different things, as the history of science teaches us. Everything of importance has been said before by somebody who did not discover it." And if the reward system of science is to function in a productive manner, its rewards must go to the true discoverers: to those who brought about a cognitive change.

My second and related answer stems from the view inherent in the sociology of science that "science is public, not private" (Merton 1968a, p. 450; see also Merton [1938] 1970, p. 219; Ziman 1968, pp. 8–10 et passim). Science is a cooperative venture of many researchers, and the function of a scientific discovery is not simply to enable an individual scientist to add to his private stock of knowledge but to stimulate a new research program on the part of colleagues in his field of inquiry, for only in that way can the full scientific potential of the discovery be efficiently exploited. But the probability that one scientist will stimulate others to devote themselves to a new research program along the lines he is working on is directly related to the extent that he himself is aware of the significance of his work. He is less likely to be able to "sell his product" to his colleagues—or at least (to shift to a normative plane of discourse) he would be unjustified in claiming credit for having sold it to them—if he himself is not convinced of its significance. And this means that his "product" must be part of his central message.

To avoid misunderstanding, let me emphasize that the distinction I am making is not that between chance and intended scientific findings. In particular, I would denote as a discovery even a chance finding, provided its significance was recognized by the researcher in question (e.g., Alexander Fleming and the discovery of penicillin).[4] Nor am I adopting a "productivity ethics" viewpoint and saying that the "marketplace of science" gives each his due: that if it has not accorded priority to some individual, that alone is evidence that the individual did not deserve it. I have no doubt that the rewards of science have not always been justly allocated. But I would not cite as examples of such injustice cases in which, after one individual has emphasized and received credit for a scientific discovery, "anticipations" of it are found in passages of earlier works by other individuals who at the time did not consider these passages

[4] Cf. in this context the interesting study by Barber and Fox (1958). Cf. also Barber (1952, pp. 267–71).

sufficiently important to incorporate into their central message in order to bring them to the attention of the profession.[5]

Conversely, I would cite as instructive examples of justice being done those "premature discoverers" who despite their emphasis on their respective central messages were at first ignored "because the time was not yet ripe,"[6] but who—when the time did "ripen"—were rewarded with recognition for their priority, and sometimes even with eponymous fame by having the discovery named after them.[7] The classic example is that of Mendel, whose 1865 work on heredity was rediscovered at the end of the century and gave rise to "Mendelism."[8] And in economics we have the well-known examples of Slutsky's ([1915] 1952) theory of consumer behavior with its "Slutsky equation," rediscovered by Hicks and Allen (1934) as well as by Henry Schultz (1935, pp. 439–40, 443)[9] and of Ramsey's ([1928] 1969) theory of optimal savings with its "Ramsey solution," in whose rediscovery after World War II Samuelson played a leading role (see the references cited in Samuelson [1965, p. 93, n. 1, and p. 94, nn. 3, 4]; see also Samuelson and Solow [1956, pp. 261 ff.]). These examples also suggest that the "premature discoverer" will subsequently be rewarded with eponymy only if his work is rediscovered not too long after the ripening of interest in the question—before the name for the (re)discovery has been preempted![10]

Let me conclude with several observations. First, the conclusions of this study are not those I expected when I began it: for the reasons indicated in the fourth paragraph of this paper, I expected to find that the discovery of the General Theory was indeed, as many have contended, an instance of independent multiple discovery. Thus, for reasons it is difficult for us today to comprehend—so thoroughly ingrained are we with the conceptual framework of the General Theory—the conceptual framework of the early 1930s had inherent in it certain habits of thought

[5] For a similar view, see Merton's (1968b, pp. 21 ff.) discussion of "adumbrations."

[6] This expression by itself is, of course, a tautology. In certain cases, however, it has been endowed with meaning by an analysis of the reasons that may have led to a work's being originally ignored; see, e.g., Barber (1961), Polanyi (1967), Stent (1972), and the paper by Sandler and Sandler referred to in n. 8 below.

[7] On the practice of eponymy with reference to scientific discoveries, see Merton (1957, pp. 298–300); see the same work for many examples of this practice.

[8] Cf. Barber (1961, pp. 544, 547–48, 551). Cf. also the forthcoming paper by Iris Sandler and Laurence Sandler on Mendel's discovery and rediscovery.

[9] Hicks and Allen became aware of Slutsky's prior discovery only after the publication of their own article; see Allen (1936, p. 120) and Hicks (1939, pp. 19, 309).

[10] This point has resulted from stimulating discussions with Laurence Sandler, to whom I am indebted. That this condition is not a sufficient one seems to be suggested by Stephen Stigler's provocative note on eponymy (1980), though there is a need to examine the cases he cites from the view of their respective central messages.

which made the process of achieving the new view of the world expounded by this theory a long and arduous one. What the exact nature of these mental obstacles was I do not fully understand. But that they did exist is clear from the record.[11]

Second, even if the writings of the Stockholm school and of Kalecki do not constitute the General Theory, they do have points in common with it, particularly in the case of Kalecki. Correspondingly, these writings provided important supplements to Keynes's exposition of the theory. Let me also observe that nothing I have said in this paper should be construed as belittlement of the individuals whose works have been discussed. To take some liberties with Merton's statement (1961, p. 366) that "great scientists will have been repeatedly involved in multiples," let me suggest that it is also the great scientists who are more likely to be involved in alleged or near multiples. Thus the fact that both Myrdal and Ohlin received the Nobel Prize in economics—the former for his work "relating economic analysis to social, demographic and institutional conditions," the latter for his contribution to international trade theory[12]—is evidence of their scientific stature. And Kalecki achieved comparable stature not only for his theoretical and empirical work on the investment cycle but also for his development of the principle of increasing risk and for his work on planning and socialist economics. So my question in this paper has been whether, in addition to these important contributions, these economists can also be credited with the independent discovery of the General Theory.

Nor do I criticize these economists for having claimed priority for such a discovery: for, like many of those who subsequently supported their claims, they may have been defining the General Theory in a way which differs from mine. Alternatively, they may have been succumbing to the all-too-human weakness of retroactively attributing greater emphasis to points which they had not originally regarded as part of their central message. After all, who among us has not on occasion felt that the profession has wrongly credited others with what we—with at most "a slight difference of emphasis"—had done before?

Similarly, to the extent that it is meaningful to speculate about such matters, I would not object unduly to the contention that in the course

---

[11] For a general discussion of such mental obstacles in scientific research, see Barber (1961). See also Kuhn's (1962, chap. 7) description of the slowness with which shifts have taken place from one "paradigm" to another. See also Fleck ([1935] 1979) for a detailed study of the influence of such "thought styles" in a case of medical research. On the other hand, see Polanyi (1967) for a rationalization of the initial resistance of science to completely new ideas.

[12] See the respective Nobel Prize citations in the *Scandinavian Journal of Economics* 76 (December 1974): 469; and 80 (no. 1, 1978): 62–63.

of time these economists would have discovered the General Theory independently—and again, this is particularly so for Kalecki. But economists more than anyone else should be aware that time has value. It is a commonplace that an essential part of the intangible capital of any society is its stock of scientific knowledge; so it is a matter of significance whether an increment to this capital (and to its subsequent yield of fruits) takes place at time t or at time $t + \theta$, with the degree of significance increasing with θ. In the long run everything would have been independently discovered by someone else: but students of Keynes need not be reminded that in the long run we are all dead (see Keynes's *Tract on Monetary Reform* [1923], p. 65).

One final observation: my conclusion that there was no multiple discovery of the General Theory has been based on a precise (some would say overprecise) definition of the innovative feature of this theory accompanied by a detailed examination of texts which was in turn guided by an emphasis on their respective central messages. The natural question which then arises is whether other alleged cases of independent multiple discovery in, say, economics would stand up under a similar examination, and the immediate answer is that at least some of them would not. Thus, Edward Chamberlin (1937, 1951) persistently and quite rightly emphasized that his theory of monopolistic competition differed from Joan Robinson's theory of imperfect competition. Similarly, William Jaffé (1976) examined the respective expositions of the marginal-utility theory by Jevons, Menger, and Walras and showed that they "were markedly different and influenced the future course of theoretical model building in fundamentally different ways" (p. 512). And a number of years ago, I showed (Patinkin 1965, suppl. notes C:2 and E:1) that there are significant differences between the respective cash-balance approaches to the quantity theory of Walras, Wicksell, and the Cambridge school.

Nor should we think that this "dehomogenization" of multiple discoveries on closer scrutiny is unique to economics or, more generally, to the social sciences. In a detailed study of various allegedly simultaneous technological discoveries, Jacob Schmookler (1966, chap. 10) showed that there were in fact important differences among them. Similarly, Yehuda Elkana (1970) demonstrated that the law of conservation of energy—long cited as a standard case of multiple discovery—was strictly speaking not one: "that in the span 1840–1860 different problems were bothering different groups of people in different places, and they came up with different answers. The answers turned out to be related, until finally in the 1860s they proved to be more than related, they turned out to be logically derivable from one another" (p. 32). And in a comparison of the non-Euclidean geometries invented by Bolyai and Lobachevsky, Petronievics has contended that "Lobachevsky had developed five of the nine salient

components of their overlapping conceptions more systematically, more fruitfully, and in more detail."[13]

Thus even from the viewpoint of my first criterion—the precise definition of the discovery—many alleged instances of multiple discovery turn out not to have been such after all. And I conjecture (with the hope that this conjecture will be tested by further case studies) that the number of multiple discoveries will be further reduced once account is taken of my second criterion—the extent to which the discovery was part of the scientist's central message, and hence the extent to which it influenced or might have influenced the path of scientific development. All of which leads me to the general conclusion that scientific research is less redundant—and, as a corollary, that the individual scientist is more important—than the by-now-familiar long lists of alleged multiple discoveries might lead us to believe.

REFERENCES

Allen, R. G. D. 1936. "Professor Slutsky's Theory of Consumers' Choice." *Review of Economic Studies* 3 (February): 120–29.
Barber, Bernard. 1952. *Science and the Social Order.* Glencoe, Ill.: Free Press.
———. 1961. "Resistance by Scientists to Scientific Discovery." *Science* 134 (September 1): 596–602. Reprinted in Barber and Hirsch, eds., 1962.
Barber, Bernard, and Renée C. Fox. 1958. "The Case of the Floppy-eared Rabbits: An Instance of Serendipity Gained and Serendipity Lost." *American Journal of Sociology* 64 (September): 128–36. Reprinted in Barber and Hirsch, eds. 1962.
Barber, Bernard, and Walter Hirsch, eds. 1962. *The Sociology of Science.* New York: Free Press.
Chamberlin, Edward H. 1937. "Monopolistic or Imperfect Competition?" *Quarterly Journal of Economics* 51 (August): 557–80.
———. 1951. "Monopolistic Competition Revisited." *Economica* 18 (November): 343–62.
Elkana, Yehuda. 1970. "The Conservation of Energy: A Case of Simultaneous Discovery?" *Archives Internationales d'Histoire des Sciences* 90–91 (January–June): 31–60.
Feiwel, George R. 1975. *The Intellectual Capital of Michal Kalecki.* Knoxville: University of Tennessee Press.
Fisher, Irving. 1911. *The Purchasing Power of Money: Its Determination and Relation to Credit Interest and Crises.* New York: Macmillan.
Fleck, Ludwik. (1935) 1979. *Genesis and Development of a Scientific Fact.* Translated by Fred Bradley and Thaddeus J. Trenn. Chicago: University of Chicago Press.
Frisch, Ragnar, and H. Holme. 1935. "The Characteristic Solutions of a Mixed Difference and Differential Equation Occurring in Economic Dynamics." *Econometrica* 3 (April): 225–39.
Gustafsson, Bo. 1973. "A Perennial of Doctrinal History: Keynes and 'The Stockholm School.' " *Economy and History* 16:114–28.

[13] Quoted from Merton's (1968*b*, p. 10) summary of Petronievics's study. Merton (1968, p. 10, n. 18) also states that in an earlier study Petronievics reached a similar conclusion with respect to the famous case of the discovery of the theory of evolution by both Charles Darwin and Alfred Russel Wallace. Merton refers to these studies in the course of his remarks on the difficulty of establishing "the degree of similarity between independently developed ideas," particularly in the social sciences with their "typically less precise formulations" (pp. 9–10).

Hansen, Alvin. 1951. *Business Cycles and National Income*. New York: Norton.

Hicks, J. R. 1939. *Value and Capital*. Oxford: Clarendon.

Hicks, J. R., and R. G. D. Allen. 1934. "A Reconsideration of the Theory of Value." *Economica* 1 (February): 52–76; (May): 196–219.

Hutchison, T. W. 1953. *A Review of Economic Doctrines 1870–1929*. Oxford: Clarendon.

———. 1968. *Economics and Economic Policy in Britain 1946–1966: Some Aspects of Their Inter-Relations*. London: Allen & Unwin.

Jaffé, William. 1976. "Menger, Jevons and Walras De-Homogenized." *Economic Inquiry* 14 (December): 511–24.

Kahn, R. F. 1931. "The Relation of Home Investment to Unemployment." *Economic Journal* 41 (June): 173–98. Reprinted in Kahn 1972.

———. 1972. *Selected Essays on Employment and Growth*. Cambridge: Cambridge University Press.

Kalecki, Michal. 1933. *Próba teorii koniunktury*. [An essay on the theory of the business cycle] Warsaw: Institute of Research on Business Cycles and Prices. Reprinted in Kalecki 1979.

———. 1935a. "Istota poprawy koniunkturalnej." *Polska Gospodarcza* [Economic Poland] 43: 1320–24. Reprinted in Kalecki 1979. Translated by Ada Kalecki under the title "The Mechanism of the Business Upswing" in Kalecki 1966.

———. 1935b. "A Macrodynamic Theory of Business Cycles." *Econometrica* 3 (July): 327–44.

———. 1936. "Parę uwag o teorii Keynesa." [Some remarks on Keynes's theory] *Ekonomista* 36 (no. 3): 18–26. Reprinted in Kalecki 1979.

———. 1966. *Studies in the Theory of Business Cycles: 1933–1939*. Translated by Ada Kalecki. Oxford: Blackwell.

———. 1979. *Kapitalizm: Koniunktura i zatrudnienie. Dzieła*. tom 1. [Capitalism: business cycles and employment. Works, vol. 1] Edited by Jerzy Osiatyński. Warsaw: Panstwowe Wydawnictwo Ekonomiczne, for Polska Akademia Nauk.

Keynes, John Maynard. 1923. *A Tract on Monetary Reform*. Reprinted in Keynes 1971–, vol. 4.

———. (1929a) 1949. "The German Transfer Problem." In *Readings in the Theory of International Trade*, selected by a committee of the American Economic Association. Philadelphia: Blakiston, for the American Economic Association.

———. 1929b. "The Reparation Problem: A Rejoinder." *Economic Journal* 39 (July): 179–82.

———. 1930. *A Treatise on Money*. Vol. 1, *The Pure Theory of Money*. Reprinted in Keynes 1971–, vol. 5.

———. 1931. *Essays in Persuasion*. Reprinted with additions in Keynes 1971–, vol. 9.

———. 1936. *The General Theory of Employment Interest and Money*. Reprinted in Keynes 1971–, vol. 7.

———. 1971–. *Collected Writings*. London: Macmillan, for the Royal Economic Society.

Keynes, John Maynard, and Hubert Henderson. 1929. *Can Lloyd George Do It? An Examination of the Liberal Pledge*. Reprinted in Keynes 1971–, 9:86–125.

Klein, Lawrence. 1951. "The Life of John Maynard Keynes." *Journal of Political Economy* 59 (October): 443–51.

Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lakatos, Imre. 1970. "Falsification and the Methodology of Scientific Research Programmes." In Lakatos and Musgrave 1970.

———. 1978. "Science and Pseudoscience." Pp. 1–7 in *The Methodology of Scientific Research Programmes: Philosophical Papers*, by Imre Lakatos, vol. 1. Edited by John Worrall and Gregory Currie. Cambridge: Cambridge University Press.

Lakatos, Imre, and Alan Musgrave, eds. 1970. *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press.

Laudan, Larry. 1977. *Progress and Its Problems*. Berkeley and Los Angeles: University of California Press.

Lindahl, Erik. 1939. *Studies in the Theory of Money and Capital.* London: Allen & Unwin. (Includes translation of sections of a 1930 monograph in Swedish on monetary policy).

Merton, Robert K. (1938) 1970. *Science, Technology and Society in Seventeenth-Century England.* In *OSIRIS: Studies on the History and Philosophy of Science, and on the History of Learning and Culture,* edited by George Sarton. New York: Fertig.

———. 1957. "Priorities in Scientific Discovery." *American Sociological Review* 22 (December): 635–59. Reprinted in Merton 1973.

———. 1961. "Singletons and Multiples in Science." *Proceedings of the American Philosophical Society* 105 (October): 470–86. Reprinted in Merton 1973.

———. 1968a. "The Matthew Effect in Science." *Science* 159 (January 5): 56–63. Reprinted in Merton 1973.

———. 1968b. *Social Theory and Social Structure.* Enlarged ed. New York: Free Press.

———. 1973. *The Sociology of Science: Theoretical and Empirical Investigations,* edited and with an introduction by Norman W. Storer. Chicago: University of Chicago Press.

Myrdal, Gunnar. 1934. *Finanspolitikens Ekonomiska Verkningar.* [The economic effects of fiscal policy] Stockholm: Kungl. Boktryckeriet.

———. (1931, 1933) 1939. *Monetary Equilibrium.* London: Hodge. (Translated with revisions from the 1933 German edition, which is a revision and enlargement of the original 1931 Swedish version.)

Ohlin, Bertil. (1929) 1949. "Transfer Difficulties, Real and Imagined." In *Readings in the Theory of International Trade,* selected by a committee of the American Economic Association. Philadelphia: Blakiston, for the American Economic Association.

———. 1933a. *Interregional and International Trade.* Cambridge, Mass.: Harvard University Press.

———. (1933b) 1978. "On the Formulation of Monetary Theory," translated by Hans J. Brems and William P. Yohe. *History of Political Economy* 10 (Fall 1978): 353–88.

———. 1934. *Penningpolitik, Offentliga Arbeten, Subventioner och Tullar Som Medel Mot Arbetslöshet.* [Monetary policy, public works, subsidies and tariff policy as remedies for unemployment] Stockholm: Kungl. Boktryckeriet.

———. (1937) 1944. "Some Notes on the Stockholm Theory of Saving and Investment." In *Readings in Business Cycle Theory,* selected by a committee of the American Economic Association. Philadelphia: Blakiston, for the American Economic Association.

———. 1981. "Stockholm and Cambridge: Four Papers on the Monetary and Employment Theory of the 1930s," edited by O. Steiger. *History of Political Economy* 13 (Summer): 189–255.

Palander, Tord. (1941) 1953. "On the Concepts and Methods of the 'Stockholm School,' " translated by R. S. Stedman. *International Economic Papers* 3:5–57.

Patinkin, Don. 1965. *Money, Interest, and Prices.* 2d ed. New York: Harper & Row.

———. 1982. *Anticipations of the General Theory? And Other Essays on Keynes.* Chicago: University of Chicago Press.

Pigou, A. C. (1917) 1951. "The Value of Money." In *Readings in Monetary Theory,* selected by a committee of the American Economic Association. Philadelphia: Blakiston, for the American Economic Association.

Polanyi, Michael. 1967. "The Growth of Science in Society." *Minerva* 4 (Summer): 533–45.

Ramsey, Frank P. (1928) 1969. "A Mathematical Theory of Saving." In *Readings in Welfare Economics,* selected by a committee of the American Economic Association. Homewood, Ill.: Irwin, for the American Economic Association.

Robinson, Joan. 1952. *The Rate of Interest: And Other Essays.* London: Macmillan.

Samuelson, Paul A. (1965) 1972. "A Catenary Turnpike Theorem involving Consumption and the Golden Rule." Pp. 486–96 in Samuelson 1972, vol. 3.

———. 1966–77. *The Collected Scientific Papers of Paul A. Samuelson.* Vols. 1, 2 (1966), edited by Joseph E. Stiglitz; vol. 3 (1972), edited by Robert C. Merton; vol. 4 (1977), edited by Hiroaki Nagatani and Kate Crowley. Cambridge, Mass.: MIT Press.

Samuelson, Paul A., and Robert M. Solow. (1956) 1966. "A Complete Capital Model involving Heterogeneous Capital Goods." Pp. 261–86 in Samuelson 1966, vol. 1.

Sandler, Iris, and Laurence Sandler. Forthcoming. "A Speculation on a Conceptual Ambiguity That Resulted in the Neglect of Mendel's Paper."

Schmookler, Jacob. 1966. *Invention and Growth*. Cambridge, Mass.: Harvard University Press.

Schultz, Henry. 1935. "Interrelations of Demand, Price, and Income." *Journal of Political Economy* 43 (August): 433–81.

Slutsky, Eugen E. (1915) 1952. "On the Theory of the Budget of the Consumer," translated by Olga Regusa in *Readings in Price Theory*, selected by a committee of the American Economic Association. Homewood, Ill.: Irwin, for the American Economic Association.

Steiger, Otto. 1978. "Prelude to the Theory of a Monetary Economy: Origins and Significance of Ohlin's 1933 Approach to Monetary Theory." *History of Political Economy* 10 (Fall): 420–46.

Stent, Gunther S. 1972. "Prematurity and Uniqueness in Scientific Discovery." *Scientific American* 227 (December): 84–93.

Stigler, Stephen M. 1980. "Stigler's Law of Eponymy." In *Science and Social Structure: A Festschrift for Robert K. Merton*, edited by Thomas F. Gieryn. *Transactions of the New York Academy of Sciences*, ser. 2, 39:147–57.

Thomas, Brinley. 1936. *Monetary Policy and Crises: A Study of Swedish Experience*. London: Routledge.

Tinbergen, Jan. 1935. "Annual Survey: Suggestions on Quantitative Business Cycle Theory." *Econometrica* 3 (July): 241–308.

Whitehead, A. N. (1917) 1974. *The Organisation of Thought: Educational and Scientific*. Westport, Conn.: Greenwood.

Wicksell, Knut. (1898) 1936. *Interest and Prices: A Study of the Causes Regulating the Value of Money*. Translated by R. F. Kahn. London: Macmillan.

Winch, Donald. 1969. *Economics and Policy: A Historical Study*. London: Hodder & Stoughton.

Ziman, John. 1968. *Public Knowledge: The Social Dimension of Science*. London: Cambridge University Press.

A Multivariate Model for Mobility Tables[1]

John A. Logan
URS-Berger

A general "logistic-multiplicative" model is developed which incorporates variables besides origin category to predict destination category in occupational mobility tables. The additional variables can be continuous or categorical. The (partial) bivariate relationship between origin and destination can be modeled with any of the existing multiplicative mobility models. In an empirical example, two main results emerge: interactions off the main diagonal become insignificant with the introduction of education and race variables; main diagonal effects for one origin category become insignificant while those for other categories do not, suggesting that different mechanisms of status transmission are at work in different strata.

Sociological analysis of mobility tables does not, as a rule, include multivariate specifications of the relationship of father's and son's occupations; such specifications are left to "status-attainment" models. Instead, many subtle and ingenious models hypothesizing systematic patterns in mobility tables have been proposed and tested, advancing the categorical analysis of this bivariate relationship to a refinement that would have been unimaginable 20 years ago (cf. Duncan 1982). However, the fact remains that the mobility table is nothing more than a bivariate table. It contains no information about the relationship of mobility to education, discrimination, market fluctuation, aspirations, or any other relevant factor. For this reason, mobility table analysis per se is inherently limited in the sociological illumination it can cast on mobility.

This inability of mobility tables to reveal the specific factors governing mobility has not led to abandonment of their study, since no convenient method is available which can express the previously mentioned hypotheses concerning mobility patterns if additional explanatory variables are introduced. These hypotheses hold a strong attraction for theoretically

inclined sociologists: their structure will not simply be set aside for the linear relationships of status attainment. In this paper, however, I will propose a model which both permits the representation of these hypotheses and allows estimation of the effects of additional variables (besides origins) which may help to explain destinations. Conceptually, the resulting models will resemble multiple regression models, but they will apply to categorical origins and destinations and will have a fully flexible arrangement of parameters to describe complex patterns of the bivariate origin-destination relationship. The models subsume as special cases all the models of Goodman (1965, 1969, 1972b, 1979), Hauser (1978, 1979), Duncan (1979), and others and can therefore represent the structures of those models while introducing the effects of other variables.[2]

The models proposed are still in need of further elaboration which cannot be accomplished here. The multiple regression–like equations need to be combined into causal structures, and the models deserve some discussion in terms of sociological theory. These topics will be considered later.

Although intended mainly as a simple example, an empirical application of the procedure to mobility among U.S. males will also establish an important point concerning existing mobility table models. Models which take only origins and destinations as their data are not sufficient bases for asserting that different strata, time periods, or social systems exhibit the "same" mobility regimes: similar estimated effect parameters may reflect different underlying factors or mechanisms.

LOGISTIC-MULTIPLICATIVE MODELS OF SOCIAL MOBILITY:
UNCONDITIONAL FORM

The multivariate model will now be formulated. To every multiplicative mobility model of past studies there correspond a number of more general "logistic-multiplicative" models. However, instead of presenting a general development appropriate to all logistic-multiplicative models, I will emphasize the special case of "Logistic Persistence" models. These models correspond to the off-diagonal Quasi-perfect Mobility models introduced in Goodman (1965).[3] That is, I will show how additional variables can be introduced to explain mobility table patterns while preserving the

---

[2] It should be noted that some existing models may be subsumed as special cases algebraically but perhaps not philosophically. For example, latent class models like those of Goodman (1974) and Clogg (1981) can be formulated with the structure presented in the Appendix, though it could be argued that introducing additional variables conflicts with the aims of those models.

[3] They also correspond to Duncan's (1979) model of "uniform association, diagonal omitted," as explained below.

structure of the Quasi-perfect Mobility model. This can be done for any existing multiplicative model, but the models corresponding to other multiplicative models will be left to the Appendix, since a general expression is less clear initially, though readily understood once the special case is explained.

Besides the Quasi-perfect Mobility model, two other existing models are needed to give a framework for developing and testing the Logistic Persistence model. (For convenience, multiplicative models will be written in their logged forms, i.e., as log-linear models, though the mobility table literature customarily uses the multiplicative forms.) The Saturated Multiplicative model of the mobility table is

$$\log \pi_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}, \tag{1}$$

where π_{ij} is the (unconditional) probability that an observation will fall in cell (i,j) of the table, and the parameters satisfy these identifying constraints:

$$\sum_i \alpha_i = \sum_j \beta_j = 0, \tag{2}$$

$$\sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0, \tag{3}$$

and

$$\sum_i \sum_j \pi_{ij} = 1. \tag{4}$$

The Perfect Mobility model, used as a standard by Glass (1954) and others, is given by equation (1) with the additional constraint that $\gamma_{ij} = 0$, for all i and j. Finally, the off-diagonal Quasi-perfect Mobility model of Goodman (1965) is again given by equation (1), but with this constraint:

$$\gamma_{ij} = \delta_{ij}\gamma_{ij}, \tag{5}$$

where δ_{ij} is Kronecker's delta:

$$\delta_{ij} = \begin{cases} 1 \text{ if } i = j, \\ 0 \text{ otherwise}. \end{cases}$$

When constraint (5) is imposed, the summation constraint (3) is relaxed, no longer being appropriate for identification. Using (5) as a constraint on (1) is a way of defining unique parameters for each cell of the main diagonal; this is an alternative to the usual procedure of "blocking out" the diagonal (Goodman 1972b, p. 670; 1979). In the Quasi-perfect Mobility model it is appropriate to call the remaining log-linear interactions (γ_{ij}, $i = j$) "persistence effects," since parameters greater than zero raise the probability of persisting in the state of origin.[4]

[4] This is in keeping with the usage in an early version of the Quasi-perfect Mobility model (Barton, David, and Fix 1962; Goodman 1964).

Nerlove and Press (1973) showed that any log-linear model can be expressed as a logistic regression model, which can then be extended to include continuous variables as well. In logistic form, the Saturated Multiplicative model is

$$\pi_{ij} = \frac{e^{\alpha_i + \beta_j + \gamma_{ij}}}{\sum_i \sum_j e^{\alpha_i + \beta_j + \gamma_{ij}}}, \tag{6}$$

where (2) and (3) are again appropriate constraints. Comparing (6) with its equivalent expression (1) shows that

$$\mu = -\log \left[\sum_i \sum_j \exp(\alpha_i + \beta_j + \gamma_{ij}) \right].$$

As before, setting $\gamma_{ij} = 0$ makes (6) the Perfect Mobility model, and imposing constraint (5) on (6) gives Quasi-perfect Mobility.

A CONDITIONAL FORMULATION

Though the equations above are equivalent to the usual (unconditional) specifications of multiplicative mobility models, it is reasonable to express the models in conditional terms, with the probabilities conditioned on the origin state (cf. Goodman 1965, pp. 583–84). This is reasonable because father's occupation at respondent's age 16 is indeed a fixed condition prior to the processes which determine son's occupation at later ages. (This is the logic which underlies the usual status attainment formulation.) The conditional probabilities p_{ij} satisfy these row summation constraints:

$$\sum_j p_{ij} = 1; \; i = 1, 2, \ldots I. \tag{7}$$

(instead of [4] above). They are found by the usual formula for conditional probabilities: $p_{ij} = Pr(j/i) = Pr(i \text{ and } j)/Pr(i) = \pi_{ij}/\pi_{i.}$, or, using equation (6),

$$p_{ij} = \frac{e^{\alpha_i + \beta_j + \gamma_{ij}}}{\sum_j e^{\alpha_i + \beta_j + \gamma_{ij}}}.$$

Dividing through by e^{α_i} gives

$$p_{ij} = \frac{e^{\beta_j + \gamma_{ij}}}{\sum_j e^{\beta_j + \gamma_{ij}}},$$

so that the origin category parameters drop out; the conditional probabilities do not depend on α_i. The expected frequencies and the values of the remaining parameters are unchanged in moving from the uncondi-

tional to the conditional form. Of course, the same constraints given above continue to specify Perfect and Quasi-perfect Mobility.

In the conditional form of Quasi-perfect Mobility, only two sets of parameters remain. The β_j represent tendencies to arrive in the various destination categories independently of the origin category i, while the γ_{ij}, $i = j$, represent the sole effects of origins on destinations in the model.

The effects of additional variables can be added to the general bivariate log-linear model if one specifies individual-level conditional probabilities p_{hij} for each individual $h = 1, 2, \ldots, N$ and makes these dependent on individual-level parameters:

$$p_{hij} = \frac{e^{\beta_{hj} + \gamma_{hij}}}{\sum_j e^{\beta_{hj} + \gamma_{hij}}} . \tag{8}$$

These individual-level log-linear parameters can in turn be made functions of an observed vector of continuous (or discrete) variables $\mathbf{x}_h$ and a set of shared parameters (Nerlove and Press 1973):

$$\beta_{hj} = \beta_j + \lambda_j' \mathbf{x}_h; j = 1, 2, \ldots, J , \tag{9}$$

$$\gamma_{hij} = \gamma_{ij} + \mu_{ij}' \mathbf{x}_h; i, j = 1, 2, \ldots, J . \tag{10}$$

Equations (9) and (10) have the form of regression equations, except that the left sides contain an individual's parameters appropriate to a given move, instead of his predicted values on some continuous variables. The β_j are constants corresponding in a sense to the intercept in a regression: while in a regression a single parameter suffices to determine the predicted value of the outcome variable when all independent variables are at their means, in this model (as in logistic regression) as many parameters, β_j, $j = 1, \ldots, J$, as categories are needed to determine the predicted probabilities of the outcome events at the means of the independent variables.[5] The elements of λ_j are like the independent variable coefficients in a regression. Here again J parameters take the place of each regression coefficient because the effects are not modeled as linear. The γ_{ij} in this general model, as in log-linear models, represent interaction effects between the row and column variables, origin and destination status. The elements of the μ_{ij} vectors represent the dependence of the extent of the origin-destination interactions on the values of other measured variables, continuous or discrete: each set of μ_{ij}, $j = 1, \ldots, J$, represents a three-way interaction of a measured variable with the origin and destination statuses. If the β_{hj} and γ_{hij} depend on different subsets of the independent

[5] However, one of the β_j is superfluous because of the linear restriction imposed by eq. (2).

variables, this condition is incorporated by setting appropriate elements of λ_j and μ_{ij} to zero.

Identification of the model is obtained by applying constraints (2) and (3) to the β_{kj}, β_j, γ_{kij}, and γ_{ij}; and also by requiring that

$$\sum_j \lambda_j = \mathbf{0} \, , \tag{11}$$

and

$$\sum_i \mu_{ij} = \sum_j \mu_{ij} = \mathbf{0} \, , \tag{12}$$

where $\mathbf{0}$ is a vector of zeros.

The model just described—equations (8), (9), and (10), with constraints (2), (3), (7), (11), and (12)—provides for logistic effects on each parameter of the saturated, conditional log-linear model. It will be called the Saturated Logistic-Multiplicative model. The Saturated Logistic-Multiplicative model contains as special cases both ordinary logistic regression and all multiplicative mobility models of the "usual form," that is, those models with parameters satisfying equations (1), (2), and (3) above. The latter category includes, for example, all the models considered in Goodman (1979) and Duncan (1979).[6] Inspection of the Saturated Logistic-Multiplicative model specifications (9) and (10) shows that setting $\lambda_j = \mu_{ij} = \mathbf{0}$ for all i and j gives the Saturated Multiplicative model. Specification of Perfect and Quasi-perfect Mobility within the Saturated Multiplicative model has been discussed above; all other multiplicative mobility models of the "usual form" are also special cases of the Saturated Multiplicative model and thus of the Saturated Logistic-Multiplicative model as well. Logistic Regression, however, is obtained from the Saturated Logistic-Multiplicative model by setting $\gamma_{ij} = 0$ and $\mu_{ij} = \mathbf{0}$, for all i and j. Logistic Regression is not a special case of the Saturated Multiplicative model but is contained in the Saturated Logistic-Multiplicative.[7]

Other constraints represent new models within the Saturated Logistic-Multiplicative model. Specifying $\mu_{ij} = \mathbf{0}$ leaves logistic regression effects on the destination parameters. This model is Logistic Regression with Saturated Multiplicative parameters.

[6] The reader may remember that constraint (3) was dropped in specifying the Quasi-perfect Mobility model. That difference appears to place the model outside the category of models of the "usual form." But this is only appearance. If, instead of (5), this equation is used to define the interactions for Quasi-perfect Mobility: $\gamma_{ij} = \delta_{ij}\gamma_i + [1 - \delta_{ij}]\kappa_i$; then constraint (3) can be applied, and the model falls within the "usual form" after all. However, some models, such as Hauser's (1978, 1979), do not in general satisfy (3).

[7] The term "Saturated Multiplicative model" in this sentence refers to the bivariate mobility table model. It does not refer to some other, more encompassing, multivariate multiplicative model.

As in the case of Quasi-perfect Mobility, Kronecker's delta can be used to define persistence effects within the Saturated Logistic-Multiplicative model. Retaining (9), but specifying

$$\gamma_{hij} = \delta_{ij}\gamma_{hij} = \delta_{ij}\gamma_{ij} + \delta_{ij}\mu_{ij}'x_h \qquad (13)$$

instead of (10), and relaxing constraints (3) and (12) give one of two Logistic Persistence models (LPb). In this model the interaction terms γ_{ij} and μ_{ij}, constrained to no effect off the main diagonal, acquire a more specific meaning. The γ_{ij} are "persistence effect" dummies: like group dummies in regression, they identify categorical variables, in this case origin categories. Unlike group dummies in regression, however, they do not shift the predicted distribution uniformly up or down the scale of the dependent variable. Their effect is instead to localize the predicted distribution, increasing the chance (for positive parameter values) that a son's destination will equal his origin. While decreasing the combined probabilities of arriving in other destinations, the parameter has no effect making one or another of them more likely. The μ_{ij} parameters are coefficients of measured variables: they allow measured variables to determine the degree of the localizing effect for individuals with particular characteristics. If it is stipulated that $\mu_{ij} = 0$, for all i and j, the model becomes one of constant persistence effects (with variable logistic regression effects)—a second, more constrained Logistic Persistence model (LPa).

The model definitions just described are summarized in table 1, which also gives the degrees of freedom (df) associated with each model. Hierarchical relationships among the models are shown in figure 1; models are special cases of those to which they connect by ascending lines. (The dotted lines connect models which are not central to the main arguments of this paper.) The numerical information in figure 1 will be discussed below.

While the Saturated Logistic-Multiplicative model is the most general one which will be considered in detail, the Appendix describes an even more general model which adds variable effects to any of the parameters in the multiplicative models of the large class described by Goodman (1972b, sec. 3). The model of the Appendix directly subsumes Hauser's (1978) models, which, unlike the Saturated Logistic-Multiplicative model, do not satisfy equation (3). However, the illustrative tests below of special cases of the Saturated Logistic-Multiplicative model could be extended to the parametric framework of Hauser's models by taking the model of the Appendix as the baseline.

AN EMPIRICAL EXAMPLE

I will now demonstrate how introducing additional variables into the analysis of a mobility table can account for patterns for which relatively

TABLE 1

SUMMARY OF MODEL DEFINITIONS

Symbol	Model	Destination Effects (β_j)	Persistence Effects ($\delta_{ij}\gamma_{ij}$)	Saturated Effects (γ_{ij})	Degrees of Freedom*
P.........	Perfect Mobility	Constant	...	...	$(J-1)(N-1)$
QO.........	Quasi-perfect Mobility	Constant	Constant	...	$(J-1)(N-1) - J$
S.........	Saturated Multiplicative	...	...	Constant	$(J-1)(N-J)$
LR.........	Logistic Regression	Variable	...	...	$(J-1)(N-K-1)$
LPa.........	Logistic Persistence A	Variable	Constant	...	$(J-1)(N-K-1) - J$
LPb.........	Logistic Persistence B	Variable	Variable	...	$(J-1)(N-K-1) - J(L+1)$
LRSM.........	Logistic Regression + Saturated Multiplicative	Variable	...	Constant	$(J-1)(N-K-J)$
SLM.........	Saturated Logistic-Multiplicative	...	...	Variable	$(J-1)[N-J(K+1)]$

* N = sample size, J = number of categories, K = number of independent variables (excluding constant) affecting either destination or saturated parameters, L = number of independent variables (excluding constant) affecting persistence.

complicated multiplicative models would otherwise be needed. To keep the exposition simple, I have chosen only two variables, one an example of a continuous variable (years of education) and the other discrete (race). I do not mean to recommend these two variables as a final specification of a mobility process; they are only examples of the form which a final model might take. Articulation of the Logistic Persistence model with a sociological theory of mobility is needed before an appropriate specification can be justified, but such an articulation will be considered in a later paper because of space limitations. The present discussion is intended to raise the issue of additional variables in mobility table models and to develop the algebra of the logistic-multiplicative models which incorporate them.

The 1972–78 merged General Social Surveys (GSS) of the National Opinion Research Center (NORC) provide the data for the illustrative application. The models presented above are applied to the intergenerational mobility of males in the labor force[8] who were between 25 and

SLM

LRSM   43.0
        (32)   25.4
                   (33)

S   259.6*
      (8)   12.0
               (11)   LPB

                        29.6
                        (10)
41.4*   LPA
(11)

Q0   289.0*
       (8)   65.4*
                   (5)

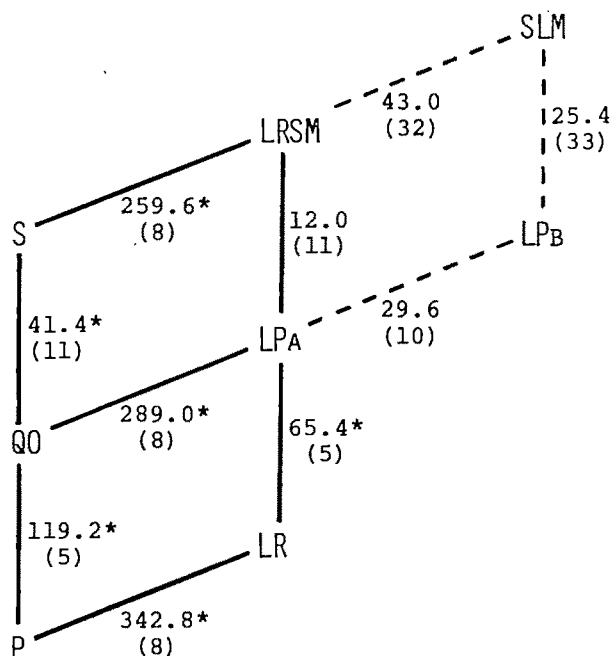119.2*   LR
 (5)

P   342.8*
          (8)

FIG. 1.—Selected mobility models applied to five-category intergenerational occupational mobility of labor-force males aged 25–34 in 1972–78. $N = 838$. (Connecting lines show likelihood-ratio χ^2 and df for test of lower model within upper. The additional variables are *EDUC* and *BLACK*. See table 1 for summary of model definitions.)

[8] "In the labor force" refers to those who reported they were "working full time," "working part time," "with a job, but not at work because of temporary illness, vacation, strike," or "unemployed, laid off, looking for work."

34 years old when interviewed during one of the seven annual GSS surveys and who had nonmissing values for father's and respondent's educations and occupations. Table 2 shows the cross-classification of respondent's current occupation by father's occupation at respondent's age 16, for the 838 individuals in the sample. Of the two variables selected as possible additional explanations of the mobility patterns in the sample, race is coded as black/nonblack, with black scored + 1 and nonblack 0; this variable is labeled *BLACK* below. Education is measured in years; it is the variable *EDUC* as found in the GSS codebook (NORC 1978).

Table 3 presents results of fitting the previously described multiplicative

TABLE 2

MOBILITY FROM FATHER'S OCCUPATION AT AGE 16 TO CURRENT OCCUPATION IN 1972–78, FOR LABOR-FORCE MALES AGED 25–34 AT TIME OF SURVEY (Frequencies)

FATHER'S OCCUPATION	SON'S OCCUPATION				
	A	B	C	D	E
A. Professional, technical, and managerial	109	36	27	24	1
B. Sales and clerical	43	8	16	15	0
C. Craftsmen and kindred workers	70	27	79	55	1
D. Operatives, service, and laborers	58	27	54	94	2
E. Farm occupations	15	7	26	29	15

SOURCE.—*General Social Surveys, 1972–1978* (NORC 1978).
NOTE.—Included males are those for whom information was reported on father's and respondent's educations and occupations. *N* = 838.

TABLE 3

DEGREES OF FREEDOM AND LOG-LIKELIHOODS FOR SELECTED MULTIPLICATIVE AND LOGISTIC-MULTIPLICATIVE MOBILITY MODELS

Symbol	Model	Degrees of Freedom	Log-Likelihood
P	Perfect Mobility	3,348	−1,178.6
QO	Quasi-perfect Mobility	3,343	−1,119.0
S	Saturated Multiplicative	3,332	−1,098.3
LR	Logistic Regression	3,340	−1,007.2
LPa	Logistic Persistence A	3,335	−974.5
LPb	Logistic Persistence B	3,325	−959.7
LRSM	Logistic Regression + Saturated Multiplicative	3,324	−968.5
SLM	Saturated Logistic-Multiplicative	3,292	−947.0

NOTE.—The models are applied to five-category intergenerational occupational mobility of labor-force males aged 25–34 in 1972–78. *N* = 838. The additional variables are *EDUC* and *BLACK*.

and logistic-multiplicative models to five-category occupational mobility of labor-force males aged 25–34 at the time of interview.[9] The Perfect Mobility, Quasi-perfect Mobility, and Saturated models are multiplicative (log-linear) models, while the rest are logistic-multiplicative models. In each of the latter, the additional variables are respondent's education (*EDUC*) and race (*BLACK*). Each line in table 3 contains the symbol and name for a particular model, the *df* of the model, and its log-likelihood.[10] Since there are 838 individuals in the sample, each with five occupational destination alternatives, there are 838 × 5 = 4,190 total *df* in the distribution. However, 838 *df* are lost by the requirement that the probabilities of arriving at all destinations for each individual sum to 1.0. This leaves 3,352 *df* for model fitting. The number of independent parameters in a given model is subtracted from 3,352 to find its *df* (cf. Amemiya 1981, pp. 1512–13). Thus the Perfect Mobility model P, with four independent parameters, has 3,348 *df*, and the main-diagonal Quasi-perfect Mobility model QO, with nine parameters, has 3,343. (Table 1 gives symbolic *df* for the general calculation.)

In the usual, tabular treatment of the multiplicative models, the *df* are calculated on the number of cells in the mobility table, so that Perfect Mobility has 25 − 5 − 4 = 16 *df*, and Quasi-perfect Mobility has 25 − 5 − 9 = 11 (remembering in both cases to subtract 5 *df* for the row summation constraints). The difference is not fundamental but only a result of switching to the individual level of analysis. Note that, although

[9] Estimation was performed with a modified version of a logistic regression program originally written by William Greene, Department of Economics, Cornell University. Most logistic regression programs are not able to incorporate the constraints of the present model, though QUAIL, by Daniel McFadden and his associates (Berkman and Associates 1979), is an exception. However, QUAIL required four times the space and three times the execution time of the modified program to perform an estimation on a sample only three-fourths the size reported here. But QUAIL is more flexible than my program. Substantially larger problems than the example reported here could be handled by the modified program with reasonable core requirements. For example, the Logistic Persistence B (LPb) model, with all possible coefficients of measured variables on destination and persistence effects, could be estimated for a problem with 5,000 cases, 17 occupational categories, and six measured variables (aside from origin) using about 480 kilobytes of core. Even larger problems can be handled by storing the data outside the core, on disk or tape or by using a virtual machine. The modified program features variable labels, automatic calculations of space requirements, and the *t*-ratios reported in table 4 and is available from me on request. The program estimates only models LPa, LPb, and logistic regression. Also available are copies of QUAIL control cards for the present models. The Logistic Regression and Saturated Logistic-Multiplicative models were estimated with an unmodified multinomial logistic regression program. To estimate the homogeneous, multiplicative models, GLIM (Nelder 1975) was used.

[10] The "log-likelihood" refers to the natural logarithm of the kernel of the likelihood function. For an argument that twice the log-likelihood should not be interpreted as an approximate χ^2 statistic in models like the present ones, see Haberman (1978, pp. 325–28). Note that no such interpretation is made here. However, twice the difference between two likelihoods will be interpreted as a χ^2, in accordance with Haberman's argument.

the total df differ, the differences in df between models are unchanged, and it is only these differences which play a role in statistical testing. (The differences in df are shown in parentheses in fig. 1.) The likelihoods of the multiplicative models (as well as the parameter estimates) are numerically equal in the aggregate and individual-level expressions. So the Quasi-perfect Mobility model of table 3, with 3,343 df, is identical to the usual model with 11 df in the tabular case; note in particular that the statistic for the test of model QO against the complete Saturated Multiplicative model, model S, is 41.4 with a difference of 11 df in either case. Placing the tabular models in the individual-level framework allows them to be evaluated directly against the exogenous variable models.

Along the paths connecting the models in figure 1 are the likelihood-ratio statistics and df for the test of each model against its next-higher alternative(s). Each test statistic is twice the difference between the log-likelihoods of two connected models and is asymptotically χ^2 distributed with df equal to the difference in df between the two models.

The six models connected by solid lines in figure 1 are of primary importance. The column of three multiplicative models on the left side of the diagram forms the rudiments of the classical tabular analysis as it has existed since Goodman's 1965 paper. First, the Perfect Mobility model P does not fit within the Saturated Multiplicative model S; that is, it does not fit the observed frequencies. Second, while off-diagonal Quasi-perfect Mobility QO fits significantly better than Perfect Mobility ($\chi^2 = 119.2$ with 5 df), it is also significantly worse than the Saturated model ($\chi^2 = 41.4$; $df = 11$). This inadequacy of the off-diagonal Quasi-perfect Mobility model has come to be expected in tables finer than 3×3.[11] So, if we look only at the left column of multiplicative models, the GSS data seem to require one or another of the many special "structural" models designed to fill just such a gap between Quasi-perfect Mobility and the Saturated model.

To the right of the multiplicative models is a parallel column of related logistic-multiplicative models. These are Logistic Regression (LR), Logistic Persistence A (LPa), and Logistic Regression plus Saturated Multiplicative effects (LRSM). Each of these models differs from its corresponding multiplicative model only by the addition of logistic regression effects of $EDUC$ and $BLACK$. That is, each differs by the presence of five nonzero parameter vectors λ_j, $j = 1, 2, \ldots, 5$. Each vector has two elements, but one vector is redundant because of the summation constraint (11), so that in each model there are $2(5 - 1) = 8$ df associated with the added parameters. These 8 df are noted on the connecting lines

[11] Note that when the GSS young male sample is categorized as white collar, blue collar, and farm, QO does fit within S, as has usually been found ($\chi^2 = 0.2$; $df = 1$).

between each multiplicative model and its more general logistic-multiplicative relative. The three χ^2 statistics on these connecting lines range from 259.6 to 342.8, meaning that each multiplicative model is strongly rejected when compared with its immediate logistic-multiplicative relative. In turn this implies, not surprisingly, that the education and race of respondents are important factors affecting categorical, intergenerational mobility.

Since the three parallel logistic-multiplicative models (LR, LPa, LRSM) must be preferred to the multiplicative models (P, QO, S), it is appropriate to ask which of the former fits best. As Perfect Mobility was a poor fit within the Saturated Multiplicative model, so its generalization, Logistic Regression, fits poorly in LRSM ($\chi^2 = 77.4$; $df = 16$). And just as Quasi-perfect Mobility was a better fit than Perfect Mobility, so is Logistic Persistence A better than Logistic Regression ($\chi^2 = 65.4$; $df = 5$). But the similarity ends at the final comparison. In the multiplicative models, Quasi-perfect Mobility did not fit well within the Saturated Multiplicative model ($\chi^2 = 41.4$; $df = 11$). The lack of fit of Quasi-perfect Mobility within the Saturated model indicated a need for more elaborate, structural effects models of off-diagonal interactions, but controlling for measured heterogeneity in education and race allows an excellent fit using only main-diagonal persistence effects ($\chi^2 = 12.0$; $df = 11$). There is no apparent gap between Logistic Persistence A and LRSM. Thus for the GSS data the need for structural effects (triangles, crossing, or uniform association parameters, etc.) disappears when additional explanatory variables are introduced.

Two models in figure 1 remain to be discussed, Logistic Persistence B (LPb) and the Saturated Logistic-Multiplicative model (SLM). The first is statistically preferable to LPa at the .001 level. Both LPb and LPa are logistic persistence models, the difference being that variable instead of constant persistence effects are allowed in LPb. Whether this evidence of variable persistence effects should be taken very seriously is doubtful. The test statistic is not as unequivocal as the other significant tests presented in the figure, especially considering the number of simultaneous tests performed. The main point to be noted is that LPb is, like LPa, a generalization of model QO, and not of some other, more complex multiplicative model. Model LPb is not an elaboration in the direction of special interaction effects; instead, it adds individual variability to the persistence effects present in model QO. Since the main purposes here can be satisfied with the simpler LPa model, the possibly better fitting LPb will not be considered further.

The most general model in figure 1, of which all other models are special cases, is the Saturated Logistic-Multiplicative model (SLM). In this model *EDUC* and *BLACK* are allowed to affect each of the parameters

of the Saturated Multiplicative model.[12] Model SLM is equivalent to separate logistic regressions performed on each row of the table, a fact which allows economical estimation and also suggests an intuitive interpretation for the model. Acceptance of model SLM would mean that the different origin categories did not share any common parameters governing the effects of origin, education, or race on any aspect of mobility. But the insignificant test statistics of SLM against LRSM (χ^2 = 43.0; df = 32) and LPb (χ^2 = 25.4; df = 33) indicate that SLM can be rejected.

I conclude that the Logistic Persistence models A and B together provide the best fit among the models tested. There is reasonable evidence that B is preferable to A, but discussion of B is outside the scope of this paper. The fit of each of these models is markedly better than that of either the Saturated Multiplicative model or the simple Logistic Regression model. No additional multiplicative constants of the mobility model type are needed to improve the fit of the Logistic Persistence model with constant persistence effects, model A. The Logistic Persistence models state that individuals of all social origins experience the same effects of education and race distributing them across destinations, though at least some origins are characterized by unexplained persistence effects which tend to retain sons with greater (or less) probability than their educations and races predict.

THE EFFECTS OF EDUCATION, RACE, AND ORIGIN CATEGORY

Table 4 shows parameter estimates for two of the models in figure 1, the Quasi-perfect Mobility model and the simpler of the two Logistic Persistence models, version A. The parameters in each model form two distinct types. Destination parameters (β_j and λ_j), whether constant or variable coefficients, sum to zero across the five destination categories; persistence parameters (γ_{ij} and μ_{ij}) do not. This difference corresponds to a difference in the interpretations of the two types. For the destination parameters it is always the arithmetic difference between two parameters which gives them meaning; the mere magnitude or sign of a parameter in isolation is meaningless since these are the result of the arbitrary identifying constraint which was imposed. However, the magnitude and sign of a persistence parameter are meaningful, since no identifying constraint subject to arbitrary respecification has been imposed. Accordingly, two kinds of t-ratios are reported in table 4. For destination parameters the t-ratios refer to the significance of the arithmetic differences between pairs of estimates (see Long 1979). For persistence parameters the t-ratios refer

[12] Recall that model LRSM, in contrast, allows *EDUC* and *BLACK* to affect only the β_{kj}, the destination parameters of logistic regression, while the saturated multiplicative interactions, the γ_{kij}, are constants.

to the significance of each estimate's difference from zero. (To save space, t-ratios are not reported for destination constants, which are of little interest.)

After noting the differences between the destination constants of the two models in table 4, consider the effects of education on destination in the Logistic Persistence model. Education differentiates upper-white-collar destination from all others, and lower-white-collar destinations from all others but farm occupations. That is, increased education significantly improves chances of an upper-white-collar occupation versus all others and significantly improves chances of a lower-white-collar occupation versus any blue-collar occupation but not versus farming. The significance of these differences is indicated by the t-ratios of differences shown below the parameter estimates for the effects of $EDUC$. All four differences between the upper-white-collar effect (.404) and the other destinations are significant, as indicated by the asterisked t-ratios in the row labeled "t vs. upper white." Three of the four differences from the lower-white-collar effect (.140) are significant: the difference from upper-white-collar

TABLE 4

PARAMETER ESTIMATES FOR TWO MODELS APPLIED TO THE FIVE-CATEGORY
INTERGENERATIONAL OCCUPATIONAL MOBILITY OF LABOR-FORCE MALES
AGED 25–34 IN 1972–78

| | DESTINATION | | | | |
MODEL AND PARAMETERS	Upper White	Lower White	Upper Blue	Lower Blue	Farm
Quasi-perfect Mobility (QO):					
Destination constants	1.05	.32	.78	.74	−2.88
Persistence constants	.91	−.58	.42	.72	3.39
	(5.2)*	(−1.5)	(2.4)*	(4.2)*	(5.9)*
Logistic-Persistence A (LPa):					
Destination constants	−4.42	−1.48	3.21	4.21	−1.52
Effects of $EDUC$ on destinations:	.40	.14	−.19	−.27	−.09
t vs. upper white	...	(5.1)*	(11.4)*	(12.3)*	(3.7)*
t vs. lower white		...	(5.7)*	(6.9)*	(1.7)
t vs. upper blue			...	(1.8)	(−.7)
t vs. lower blue				...	(−1.3)
Effects of $BLACK$ on destinations:	−.46	.02	.44	.92	−.93
t vs. upper white	...	(−.9)	(−2.0)*	(−3.3)*	(.4)
t vs. lower white		...	(−.9)	(−2.0)*	(.8)
t vs. upper blue			...	(−1.6)	(1.3)
t vs. lower blue				...	(1.7)
Persistence constants	.16	−.52	.48	.46	3.43
	(.8)	(−1.3)	(2.5)*	(2.4)*	(5.8)*

NOTE.—t-ratios are in parentheses. For persistence constants, t-ratios are given for the hypothesis that the parameter is equal to zero. For destination effects, the hypothesis is that two such effects are equal to each other.
* $t \geq 2.0$.

effect in the "t vs. upper white" row, and the differences from both blue-collar effects in the "t vs. lower white" row. However, the difference between the lower-white-collar effect and the farm effect of education (namely, .140 − [−.093] = .233) is not significant; education does not significantly distribute sons preferentially into lower-white-collar versus farm occupations. Finally, years of formal education shows no significant differentiating effects among upper-blue-collar, lower-blue-collar, and farm occupations.

These varied findings about the effects of education demonstrate a potentially important advantage of the present logistic-multiplicative method over status attainment's linear regression. In linear regression the effect of education is unitary and either present or absent—implicitly, an education effect raises destination status expectations all along the linear status scale. Another year of education should be worth so many extra status points no matter what the predicted region of the destination status scale. But here there is the suggestion of a differentiated education effect: education may raise the probability of a white-collar occupation, but if such a job is not obtained, then formal education may be irrelevant to chances of an upper- instead of lower-blue-collar occupation.

As for the variable *BLACK*, the three significant t-ratios suggest that being black makes any blue-collar occupation more likely than an upper-white-collar job, and a lower-blue-collar job more likely than a lower-white-collar job, even after controlling for education and the significant persistence effects applying to all blue-collar sons.

Farm destinations are not well differentiated by respondent's education or race: the only significant t-ratio is the one versus the upper-white-collar effect for education. This lack of discrimination is not due solely to a dearth of farm sons, for table 1 shows there are more farm sons (92) than lower-white-collar sons (82). Other independent variables than education and race may be needed to differentiate sons likely to enter farming.

The persistence effects of model LPa are significantly positive in both blue-collar categories and strikingly strong in the farm occupations. Even controlling for the influences of education and race on destination, farm sons are much more likely to wind up on the farm than chance would predict. However, there is no evidence that white-collar origins make a son more likely to find white-collar employment, apart from the relationship of education and race to such origins. There is a marked difference between the Logistic Persistence and Quasi-perfect Mobility models in the magnitude and significance of the persistence effect in the upper-white-collar category. In the Quasi-perfect Mobility model this effect is 0.91, five times its standard error, while in the Logistic Persistence model it is reduced to 0.16 and statistical insignificance. In other words, the apparent persistence effect (or status inheritance) in the highest occupa-

339

tional category—which is routinely found in bivariate mobility table anal-
ysis—is explained by the educational and racial characteristics of the
individuals originating there. In the manual and farm occupations, how-
ever, a persistence effect remains even after controlling for education and
race. The results suggest there is no single explanation of status persis-
tence: while education and race account for upper-white-collar persis-
tence, there appear to be other, unidentified factors which make blue-
collar and farm sons more likely to be immobile than their educations
and races would predict.

CAUSAL INTERPRETATIONS

Equations (8), (9), and (10), taken together, constitute a logistic regression
model which corresponds directly to a multiple regression model for a
continuous dependent variable (cf. Nerlove and Press 1973). Just as in a
multiple regression, there is no inherent ordering of temporal or causal
precedence among the independent variables, including in this case the
origin category. In the example, education would be considered temporally
(and causally) posterior to the origin category, while race might be con-
sidered a simple covariate of origin. It is possible to construct path models
specifying the causal relationships among the variables introduced to
explain the mobility table patterns, but this important refinement will
not be developed here. In the absence of such models, the interpretation
of the present model as analogous to a regression equation incorporating
origin category should be adequate. One important point which is trans-
ferable from multiple regression is that all variables having significant
direct paths to the ultimate effect in a recursive path model should have
significant coefficients in the simple regression of that effect on all causes
(cf., e.g., Duncan 1970). In an appropriate path model, the disappearance
of the persistence effect in the upper-white-collar category in the Logistic
Persistence model (table 4) suggests that no unique direct effect of upper-
white-collar origins on upper-white-collar destinations exists: the effect
would be mediated through education and possibly shared with race.
However, direct paths exist in the case of the three lower categories. This
is evidence that similar multiplicative mobility model effects may reflect
different underlying mechanisms.

EXPECTED FREQUENCIES

In logistic-multiplicative models, goodness of fit is judged on the indi-
vidual level; goodness of fit to the social mobility table per se is a secondary
question, not relevant to the optimal tests of the models presented above.
Nevertheless, it is true that the proposed models reproduce the social

mobility table (table 1) better than their tabular counterparts, and examining the expected frequencies implied by the different models gives a qualitative indication of the effects of the parameters of various models which cannot be obtained by examining χ^2 statistics. Table 5 displays the observed row proportions corresponding to the frequencies from table 2 together with the expected proportions under Perfect Mobility, off-diagonal Quasi-perfect Mobility, Logistic Regression, and Logistic Persistence A with constant persistence effects.

The Logistic Persistence model can be considered the combination of Quasi-perfect Mobility with Logistic Regression, and the degree to which

TABLE 5

OBSERVED ROW PROPORTIONS AND PROPORTIONS PREDICTED
UNDER FOUR MOBILITY MODELS

	DESTINATION				
MODEL AND ORIGIN	Upper White	Lower White	Upper Blue	Lower Blue	Farm
A. Observed row proportions:					
Upper white	.55	.18	.14	.12	.005
Lower white	.52	.10	.20	.18	0
Upper blue	.30	.12	.34	.24	.004
Lower blue	.25	.11	.23	.40	.009
Farm	.16	.08	.28	.32	.16
B. Perfect Mobility:					
Upper white	.35	.12	.24	.26	.02
Lower white	.35	.12	.24	.26	.02
Upper blue	.35	.12	.24	.26	.02
Lower blue	.35	.12	.24	.26	.02
Farm	.35	.12	.24	.26	.02
C. Quasi-perfect Mobility:					
Upper white	.55*	.11	.17	.16	.005
Lower white	.35	.10*	.27	.27	.007
Upper blue	.29	.14	.34*	.22	.006
Lower blue	.26	.13	.20	.40*	.005
Farm	.28	.14	.22	.21	.16*
D. Logistic Regression:					
Upper white	.52	.14	.17	.16	.02
Lower white	.43	.13	.22	.21	.02
Upper blue	.33	.13	.26	.27	.03
Lower blue	.26	.12	.27	.32	.03
Farm	.23	.11	.29	.35	.03
E. Logistic Persistence (A):					
Upper white	.55*	.15	.15	.14	.005
Lower white	.46	.10*	.22	.22	.006
Upper blue	.31	.13	.34*	.22	.006
Lower blue	.25	.12	.23	.40*	.006
Farm	.21	.11	.24	.28	.16*

* Identical to the observed value in this model.

its empirical fit is due to each component can be observed in the improvement each offers over Perfect Mobility, the simplest mobility model, in predicting the observed frequencies. Comparing panels A and B of table 5 shows the poor fit of Perfect Mobility, in which all row proportions are identical. All main-diagonal entries are less than the observed values except for the lower-white-collar category. Both Quasi-perfect Mobility (panel C) and Logistic Regression (panel D) improve on Perfect Mobility. On the main diagonal, Quasi-perfect Mobility fits the observations exactly, by construction. Logistic Regression improves on Perfect Mobility in the four categories other than lower white collar and does so not by construction but by taking account of the education and race variables (while ignoring origin categories). In the two diagonals immediately off the main, corresponding to movement up or down one category, Quasi-perfect Mobility improves the fit over Perfect Mobility in one cell of eight, while Logistic Regression does so in four of eight. Each component contributes to the fit of the Logistic Persistence model (panel E), in which the main diagonal fits exactly, by construction, and improvement is made over Perfect Mobility in six of the eight off-main-diagonal elements. The improvement in fit in the two off-main diagonals compared with both Perfect and Quasi-perfect Mobility reflects the fact that adjacent origin categories are producing sons relatively similar in education and race.

COMPARISONS WITH OTHER, RELATED MODELS

Though the models of this paper close the gap between multiplicative mobility models and logistic regression models in a clearly defined way, it may still be helpful to compare them with other models which have had similar aims.[13]

Spilerman (1972) used dummy-variable regression to estimate exogenous variable effects on the transition probabilities of a Markov chain. Markov transition probabilities are formally just conditional probabilities like those in equation (7) above, so the conditional logistic multiplicative models can be considered exogenous variable Markov models in appropriate contexts. Though Spilerman used separate regression equations for

---

[13] McCann (1977) proposed a model which attributes the general pattern of the mobility table to an underlying, unobserved "quality" dimension of individuals. However, the model gives no indication of the particular characteristics of individuals which determine their relative "quality" as they participate in a mobility process. Although its emphasis (in principle) on individual differences is a similarity with Logistic-Persistence and Regression models, its particular structure corresponds to a theoretical explanation from economics (Thurow 1969, 1975) which the present models do not incorporate. Therefore, McCann's model will not be given formal treatment here. Breiger's (1981) method for determining combinations of occupational categories which satisfy a definition of social class is not an alternative approach to the one of this paper so much as a step which could be performed prior to fitting the present models.

each cell of the transition matrix, he noted that the summation constraint (7) made one such equation in each row superfluous and in a footnote mentioned the possibility of using logistic regression (on each row), which would take the constraint into account (p. 289, n. 18). Had he used logistic regression, he would have estimated the Saturated Logistic-Multiplicative model of this paper. So Spilerman was in effect comparing that model with the Saturated Multiplicative model, the conditional form of the latter model being equivalent to the homogeneous Markov chain's transition matrix $P = \{p_{ij}\}$. Recall that here the Saturated Logistic-Multiplicative model was rejected in favor of the Logistic Persistence models. Logistic Persistence models, when they fit, are preferable to the Saturated Logistic-Multiplicative model because of their relative parsimony. Of course, Spilerman was not concerned with evaluating various multiplicative models contained in the Saturated Multiplicative model for their plausibility when controlling for heterogeneity, and his method was unsuitable for such a task.

Duncan (1979) described models of "uniform association" and "row effects" in off-diagonal cells of the mobility table. Though he and, especially, Goodman (1979) developed the models in such a way as to make clear that they were multiplicative mobility models of Goodman's (1972b) type, Duncan claimed the models combined features of the regression and contingency table approaches. Thus the parameter of uniform association, b, was called an analog of the (linear) regression coefficient, though it was "not, of course, a·regression coefficient in the usual sense of the word" (1979, p. 797). It is possible to reinforce in a somewhat surprising way Duncan's claim to combining regression and contingency table features within the perspective of logistic-multiplicative models. While Goodman (1979) clearly showed that Duncan's models can lie on the line between Quasi-perfect Mobility and the Saturated Multiplicative model represented here in figure 1, it is equally possible to place them on the line between the former model and Logistic Persistence A so that they become special cases of the Logistic Persistence model. In terms of equations (9) and (10) above, Goodman's positioning involves setting the λ_j and μ_{ij} terms to zero, and constraining β_j and γ_{ij} differently than for Quasi-perfect Mobility.[14] In contrast, an alternative development of Duncan's model places the Quasi-perfect Mobility constraints on β_j and γ_{ij} and then utilizes the logistic regression destination effect parameters λ_j to arrive at Duncan's b. This development shows that Duncan's b is a difference between logistic regression effects, which is tantamount to its being a regression coefficient, since it is only differences between logistic regression coeffi-

[14] More precisely, α_i would also be present since Goodman implicitly operates with an unconditional model (though Duncan's interpretation is conditional). The difference is unimportant here.

cients which have any meaning, as explained above. However, unnatural-seeming constraints are required to produce logistic-persistence models equivalent to Duncan's models. Both the fact that they can be expressed as logistic persistence models and the particular constraints needed so to express them provide an interesting interpretation for Duncan's models. However, to show the movement from Duncan's odds-ratio specifications to the parametric logistic persistence specification takes much more space than is available; interested readers are invited to write for the details.

It would be possible to approximate the models in this paper by using log-linear models which included categorical measures of variables like education instead of continuous measures.[15] Conceptually, the results would be very similar. Practically, some disadvantages would result. The number of parameters characterizing the effect of a categorized variable would be much greater than the number for its continuous counterpart. In the example above, there were five parameters describing the distributing effects of education on the five destinations; if education were trichotomized, the number of effects would be 15. This number would continue to multiply as finer education categories were used in order to reduce the loss of information inherent in categorization. The number of parameters to be estimated and interpreted would also increase roughly geometrically with the addition of further explanatory variables. Some log-linear programs (e.g., GLIM [Nelder 1975]) would, however, allow constraints to be placed on parameters, nearly eliminating this disadvantage. In general, using a log-linear model instead of a logistic-multiplicative model is comparable to using analysis of variance instead of regression in the linear case.

SUMMARY

Investigation of individual differences on characteristics other than origin category has shown that a simple persistence model fits five-category mobility data when its homogeneous variant (main-diagonal-deleted Quasi-perfect Mobility) does not. The persistence effects in the model are less important in explaining the mobility patterns than are the simple logistic regression effects the other variables exert on destinations independently of origins. Viewing the total likelihood (i.e., 218.9) between the Perfect Mobility model and the best-fitting Logistic Persistence B model as an analog of explained variance (see Goodman 1972a), 78% (100 × 171.4/ 218.9) of the "variance" is explained by the logistic effects of variables

[15] Reviewers brought to my attention a paper, at that time unpublished, by Yamaguchi (1983) which carries out such an analysis.

other than origin, leaving only 22% explained by persistence, including variable persistence effects.

That a model (LRSM) adding all possible constant log-linear effects to the effects of education and race was not significantly better than the model adding only persistence effects (LPa) means that there is no evident need for any of the special multiplicative models proposed in recent years to explain the patterns beyond simple status persistence in these data. It seems possible that to some degree these multiplicative models have been dealing with patterns which may vanish under a suitable specification of additional measured variables (other than origin). Status persistence confined to single categories has a straightforward interpretation, unlike the effects in some more elaborate models. The approach of this paper suggests that what is most needed for a better understanding of categorical status transmission is not more elaboration of new mobility effects but sustained empirical research into the specific additional factors that can best account for mobility patterns without reference to a persistence effect, and the specific factors that magnify or diminish whatever residual persistence effects remain after accounting for the first factors. It seems very likely that the persistence effects remaining in the models estimated here can be attributed in part to uncontrolled factors of the type examined in the status-attainment literature; this needs to be determined, using the methods outlined.[16]

As a step toward fulfillment of the project of explaining occupational mobility with additional measured variables, the empirical application showed mainly the proposed procedure, without much sociological content. A more empirically oriented application was omitted because space prevented including two necessary further developments: a grounding of the model in sociological theory and a causal, path-analytic treatment. However, the example did serve to emphasize a drawback of multiplicative models which take as data only origins and destinations. The persistence effects in the Quasi-perfect Mobility model behaved differently among the origins when education and race were introduced. The effect for the upper-white-collar category disappeared (table 4), while the effects for the three lowest categories remained significant. This strongly suggests that white-collar status persistence occurs via an education mechanism (though race must be partialled out in a true path model) while persistence in the other categories depends heavily on other, unspecified factors. That such different mechanisms may underlie "structurally" identical effect

[16] It is not suggested that the other multiplicative effects mentioned will never prove to be important after the introduction of other relevant variables besides origin. The Appendix provides a way of estimating such effects in the presence of other variables if they are needed.

parameters in different origins or, by implication, different periods or social systems means that one must view with great caution any conclusion that two origins, periods, or systems are "similar" in their mobility regimes simply because they show similar multiplicative mobility model patterns. This point undermines one main claim for the usefulness of such models: that they excel in intertemporal and cross-cultural comparisons. Simultaneously, the finding that education and race can account only for upper-white-collar status inheritance suggests the need for further empirical studies to attempt to replicate this result and to find the factors and mechanisms involved in status persistence among manual and farm sons.

APPENDIX

This Appendix describes the general, intrinsically unrestricted logistic-multiplicative model. Here a capital letter will usually denote a matrix or vector, and a lowercase letter a scalar.

The general formulation extends the results of Nerlove and Press (1973) to the general parametric structure of Goodman (1972b). Since virtually all multiplicative mobility models can be expressed in the terms of Goodman's paper, the formulation below allows, in a wide variety of contexts, the incorporation of continuous (or categorical) regression-like effects, and the evaluation of hypotheses holding that uncontrolled individual-level heterogeneity explains apparent multiplicative effects.

Goodman (1972b, pp. 663–65) formulates the general, intrinsically unrestricted multiplicative model in terms of a set Λ of parameters λ_w, $w = 1, \ldots, W$. Here the logged parameters, $\phi_w = \log \lambda_w$, are more convenient. The total set of constant parameters is $\Phi = \{\phi_1, \phi_2, \ldots, \phi_W\}$, while Φ_{ij} denotes the subset of constant parameters which apply to cell (i,j) of the table. (As in the models of the text, Φ will not contain a grand mean term appearing in all cells.) Then the delta defined by Goodman is

$$\delta_{ij}^w = \begin{cases} 1 \text{ if } \phi_w \in \Phi_{ij} , \\ 0 \text{ otherwise} . \end{cases}$$

If the deltas are arrayed in vectors $\Delta'_{ij} = [\delta_{ij}^1 \ \delta_{ij}^2 \ \ldots \ \delta_{ij}^W]$ and if the total parameter set Φ is considered as a corresponding vector of length W, then Goodman's general multiplicative model in logistic form is

$$\pi_{ij} = \frac{e^{\Delta'_{ij}\Phi}}{\sum_i \sum_j e^{\Delta'_{ij}\Phi}} .$$

For each individual $h = 1, \ldots, N$, a vector of V additional, measured

variables is available: $X'_k = [x_1 \, x_2 \, \ldots \, x_V]$. Corresponding to each constant effect ϕ_w is a vector of coefficients of the additional variables, $\Psi'_w = [\psi_{w1} \, \psi_{w2} \, \ldots \, \psi_{wV}]$, any number of the elements of which may be set to zero a priori.

If the W coefficient vectors are collected in a matrix Ψ of size $W \times V$, $\Psi' = [\Psi_1 \, \Psi_2 \, \ldots \, \Psi_W]$, then the general, intrinsically unrestricted logistic multiplicative model is

$$\pi_{kij} = \frac{e^{\Delta'_{ij}(\Phi + \Psi X_k)}}{\displaystyle\sum_i \sum_j e^{\Delta'_{ij}(\Phi + \Psi X_k)}}, \tag{A1}$$

which has the form of a constrained logistic regression. The conditional model for the p_{kij} looks the same, except that the denominator sum is only over j. (As in the text, the conditional model may have fewer parameters than the unconditional.)

The general model (A1) requires identifying constraints before particular models can be estimated. The constraints are parallel to the kinds discussed in Goodman (1972b). Also as in multiplicative models, apparently different parametric structures may be equivalent models (see Goodman 1979). Since (A1) is a logistic regression model, standard estimation methods apply, with appropriate constraints.

As an example of the use of equation (A1) to represent particular models, the specifications shown in figure A1 produce the conditional Logistic Persistence A model, discussed in the text. The necessary identifying constraints for this model are also discussed in the text.

Formula (A1) is also appropriate for rectangular (but not square) cross-classifications. By adding extra subscripts (as π_{hijk}), the formula becomes appropriate for logistic-multiplicative models of multiway (instead of two-way) cross-classifications with additional variables.

$$\Delta_{ij} = \begin{bmatrix} \delta_{1j} \\ \cdot \\ \cdot \\ \cdot \\ \delta_{Jj} \\ \delta_{ij}\delta_{1j} \\ \cdot \\ \cdot \\ \cdot \\ \delta_{ij}\delta_{Jj} \end{bmatrix}_{2J \times 1} \quad \Phi = \begin{bmatrix} \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_J \\ \gamma_1 \\ \cdot \\ \cdot \\ \cdot \\ \gamma_J \end{bmatrix}_{2J \times 1} \quad \Psi = \begin{bmatrix} \lambda_{11} & \cdots & \lambda_{1V} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \lambda_{J1} & \cdots & \lambda_{JV} \\ 0 & & 0 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ 0 & \cdots & 0 \end{bmatrix}_{2J \times V}$$

FIG. A1

347

American Journal of Sociology

REFERENCES

Amemiya, Takeshi. 1981. "Qualitative Response Models: A Survey." *Journal of Economic Literature* 14:1483–1536.
Barton, D. E., F. N. David, and E. Fix. 1962. "Persistence in a Chain of Multiple Events." *Biometrika* 49:351–57.
Berkman, Jerry, and Associates. 1979. "QUAIL 4.0 User's Manual." Mimeographed. Berkeley: Department of Economics, University of California.
Breiger, Ronald L. 1981. "The Social Class Structure of Occupational Mobility." *American Journal of Sociology* 87:578–611.
Clogg, Clifford C. 1981. "Latent Structure Models of Mobility." *American Journal of Sociology* 86:836–68.
Duncan, Otis Dudley. 1970. "Partials, Partitions and Paths." Pp. 38–47 in *Sociological Methodology 1970*, edited by Edgar F. Borgatta and George W. Bohrnstedt. San Francisco: Jossey-Bass.
———. 1979. "How Destination Depends on Origin in the Occupational Mobility Table." *American Journal of Sociology* 84:793–803.
———. 1982. "Review Essay: Statistical Methods for Categorical Data." *American Journal of Sociology* 87:957–64.
Glass, D. V., ed. 1954. *Social Mobility in Britain*. Glencoe, Ill.: Free Press.
Goodman, Leo A. 1964. "The Analysis of Persistence in a Chain of Multiple Events." *Biometrika* 51:405–11.
———. 1965. "On the Statistical Analysis of Mobility Tables." *American Journal of Sociology* 70:564–85.
———. 1969. "On the Measurement of Social Mobility: An Index of Status Persistence." *American Sociological Review* 34:831–50.
———. 1972a. "A General Model for the Analysis of Surveys." *American Journal of Sociology* 77:1035–86.
———. 1972b. "Some Multiplicative Models for the Analysis of Cross-Classified Data." Pp. 649–96 in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*. Vol. 1. Edited by L. Le Cam et al. Berkeley and Los Angeles: University of California Press.
———. 1974. "The Analysis of Systems of Qualitative Variables When Some of the Variables Are Unobservable. Part I—a Modified Latent Structure Approach." *American Journal of Sociology* 79:1179–1259.
———. 1979. "Multiplicative Models for the Analysis of Occupational Mobility Tables and Other Kinds of Cross-Classification Tables." *American Journal of Sociology* 84:804–19.
Haberman, Shelby J. 1978. *Analysis of Qualitative Data*. Vol. 1. New York: Academic Press.
Hauser, Robert M. 1978. "A Structural Model of the Mobility Table." *Social Forces* 56:919–53.
———. 1979. "Some Exploratory Methods for Modeling Mobility Tables and Other Cross-classified Data." Pp. 413–58 in *Sociological Methodology 1980*, edited by Karl F. Schuessler. San Francisco: Jossey-Bass.
Long, J. Scott. 1979. "The Use and Interpretation of Multinomial Logit Analysis." Mimeographed. Pullman: Washington State University.
McCann, James C. 1977. "A Theoretical Model for the Interpretation of Tables of Social Mobility." *American Sociological Review* 42:74–90.
National Opinion Research Center (NORC). 1978. *General Social Surveys, 1972–1978: Cumulative Codebook*. Chicago: University of Chicago.
Nelder, J. A. 1975. *GLIM Manual*. Release 2. Oxford: Numerical Algorithms Group, Royal Statistical Society.
Nerlove, Marc, and S. James Press. 1973. *Univariate and Multivariate Log-Linear and Logistic Models*. Technical Report R-1306-EDA/NIH. Santa Monica, Calif.: Rand.
Spilerman, Seymour. 1972. "The Analysis of Mobility Processes by the Introduction of Independent Variables into a Markov Chain." *American Sociological Review* 37:277–94.

Thurow, Lester C. 1969. *Poverty and Discrimination.* Washington, D.C.: Brookings Institution.

————. 1975. *Generating Inequality: Mechanisms of Distribution in the U.S. Economy.* New York: Basic.

Yamaguchi, Kazuo. 1983. "The Structure of Intergenerational Occupational Mobility: Generality and Specificity in Resources, Channels and Barriers." *American Journal of Sociology* 88:718–45.

Economic Development, Pension Policies, and the Labor Force Participation of Aged Males: A Cross-national, Longitudinal Approach[1]

Fred C. Pampel and Jane A. Weiss
University of Iowa

The decline in labor force participation of aged males experienced by all developed countries can be seen as the result of two kinds of change: (1) economic growth and the need for a young, recently trained labor force has affected the educational system and occupational structure in ways that give young workers a competitive advantage and force older workers out of the labor force at mandatory retirement ages, and (2) government policies have been developed to make retirement more attractive financially and to induce older workers to leave the labor force voluntarily. This paper examines the additive and interactive influence of these two groups of variables using a sample of six time points (every five years from 1950 to 1975) for 18 relatively developed nations. Models are estimated by using generalized least squares which relate labor force participation of males aged 65 and over to such variables as the size of the agricultural labor force, the old-age dependency ratio, government revenues, social insurance and pension benefits, coverage of the population by pension systems, and the existence of means and retirement tests. The results show dominant effects from the nonpolicy variables that relate to occupational changes in industrial economies and secondary effects from policy-manipulable variables such as pension expenditures that can be changed to increase labor force participation rates and ease the strain on social security systems.

While almost all countries undergoing economic development have experienced a decline in the labor force participation rates of aged males, the largest drop has occurred in the most developed countries where retirement has become fully institutionalized (Durand 1975; Sheppard

1976). The structural determinants of this trend can be divided into two groups: (1) changes in the educational system, occupational structure, and population brought about by continued economic growth; and (2) changes in government structures and policies that involve social welfare and pension programs. The first group involves the increased demand of industrial economies for a younger labor force and the loss of choice in the retirement decisions of aged workers (i.e., push factors). The second group involves the reduction in the supply of older workers that results from attempts to attract them into retirement by providing higher pension inducements to not work (i.e., pull factors). The first group of variables involves the independent power of economic growth to force societal changes in labor force participation, while the second group relates to economic growth but is more subject to government manipulation.

Theories of industrialization emphasize the importance of the first set of demand variables. With continued economic growth, patterns of labor force participation—as well as other social institutions and aspects of work—must change in response to the needs of industrial economies (Kerr et al. 1960; Form 1979). Bureaucratized organizations in industrial economies demand highly educated workers, high turnover to bring in younger, newer talent, and efficient, productive workers with the latest skills and knowledge. This places older workers at a competitive disadvantage with younger workers. Mandatory retirement, long-term unemployment, and obsolescence of jobs of older workers force them out of the labor force, and the participation rate at the older ages declines (Burgess 1960; Cowgill 1974; Graebner 1980). This argument further predicts convergence in patterns of participation of older males at the highest levels of economic development. Traditional social, political, and economic differences among nations, which result in variations in participation rates, are dominated by the industrializing economy. Hence, not only is the mean participation rate of older males lower in highly industrialized nations, but the standard deviation of the rates is also low (Durand 1975).

An alternative theory emphasizes the importance of political and cultural differences in determining worker supply and focuses on the second set of variables described above. Because of differences in culture, political structures, and the nature of class relations, nations at the same level of economic development may provide different levels of resources for transfer, welfare, and retirement programs (Griffin, Devine, and Wallace 1981). The higher the level of transfer benefits, the more incentive for workers to retire and the lower the participation rate of older males. Because social welfare and pension expenditures of governments are at least partially independent of economic development, this argument predicts that divergencies in patterns of participation will be maintained by such government programs.

351

A third explanation focuses on the combination of the forces of industrial convergence and the growth of social welfare benefits. Initial decreases in participation can be seen as the results of the needs of industry for a younger labor force and the implementation of mandatory retirement policies. However, once nations have reached high levels of economic development, retirement has become common, and there is lengthy experience with pension systems, mandatory retirement will be less important, and induced retirement will be more important. Participation will therefore respond primarily to levels of pension and social welfare benefits. This explanation combines the previous two by suggesting that the needs of industry for older workers to retire are eventually satisfied by changes in the norms of retirement and government programs rather than by forced withdrawal from the labor force. In sum, the effect of economic growth and social policy variables on participation may differ depending on the time period, state of economic development, and experience with public pension programs.

The theories have different implications for how nations can deal with the problems of an overburdened social security system. Many nations are finding that transfer programs based on the contributions of workers have not been able to support the large number of aged persons outside of the labor force (Fisher 1978). It may become necessary to encourage workers to remain in the labor force as long as possible. If growth of industrial work structures is responsible for the decline in participation, then major changes in participation may be difficult to implement without changes in the nature of industrial labor demands. If the decline is determined more strongly by government programs, then policies can be changed to reverse the trend in participation and ease the strain on the pension systems.

The effect of these different types of structural variables on participation rates has not been studied much in the literature on retirement. Most economic and sociological research on aging has been based on individual-level, cross-sectional data and has not been able to focus on structural variation in economic growth and public policies. There have been a limited number of macro level studies of retirement, but none has used a research design that provides maximum variation in the structural variables or allows the separation of the influence of a large number of variables. For example, the most common macro level approach has been to analyze time-series data on participation for the United States (e.g., Bowen and Finegan 1969; Pampel 1981). However, many policy variables have been constant over the time period studied, and their effect on participation cannot be examined. Other studies have examined small samples of nations at one time point, thereby concentrating on cross-national variation but not on temporal variation (Cowgill and Holmes

1972; Palmore and Manton 1974; Cohn 1982). Moreover, these studies have been limited to ethnographic comparisons of societies or to analyses of bivariate relationships and have not been capable of determining the impact of a number of different structural variables simultaneously.

Our objective in this article is to test the structural explanations of the decline in participation of aged males in a manner that improves on existing studies. We take a macro, comparative, longitudinal, and multivariate approach to the problem. Aggregate indicators of the population, economy, occupational structure, government, labor force, and pension system are gathered for a sample of nations over a 25-year time span. By combining time-series with comparative data, we obtain greater variation in both the independent and dependent variables. This allows the multivariate analysis of the determinants of participation and comparison of the relative influence of industrial demand variables and worker supply, policy variables. Comparison of the relative effects will thereby provide a test of industrialism and social policy explanations of change in labor force participation patterns.

The theoretical models to be specified and tested here will be applied to nations that are relatively developed. We are primarily concerned with accounting for variations in the participation levels among countries with relatively developed institutions of retirement. If countries are not sufficiently developed to support an expensive pension system for a large nonworking population, then policy variations will be small and can have little effect on retirement (Cutright 1967; Wilensky 1975).[2] Since developing nations have not reached levels of economic growth needed to support an expensive pension system, the processes that affect participation will be different from those of developed nations, and a model including both types of nations may be misspecified. Our approach is further necessitated by the lack of reliable data for developing nations over a long time span; a smaller sample of developed nations allows us to use a dynamic research design that covers a 25-year time span. Thus, the sample consists of 18 relatively developed nations with six (every five years from 1950 to 1975) time points each.

The model and analysis in this paper will focus on labor force participation rather than on retirement. Although these concepts are closely

[2] The use of developed nations truncates the range of values of economic development and is likely to reduce the relationship between participation and development that would exist for a sample of all nations. This provides a conservative test of the effects of economic growth variables and maximizes the importance of state policy variables. As argued above, however, this is theoretically appropriate for the issues to be examined. Moreover, despite the limited sample, much variation remains in economic growth variables because of the use of data for the six time points covering 25 years (see table 1). We will use economic growth and development, then, to refer to levels in more developed nations over the past several decades.

related, retirement is more difficult to operationalize at the aggregate level. Measurements of retirement must consider the previous years worked, receipt of a pension, intentions to return to the labor force, and reduction of hours worked. Labor force participation involves the percentage of aged males currently employed or actively seeking work and can be measured without reference to subjective intentions and work histories. Because evidence in the United States has shown that measures of retirement and labor force participation have changed nearly identically over the past 30 years (Pampel 1981), and the theoretical determinants of one affect the other, we will use the more easily measured macro concept of labor force participation in the theory and analyses that follow.

MODEL SPECIFICATION

The industrialization and social policy explanations of the decline in participation of older males in developed countries are not mutually exclusive. Each may contribute to the understanding of the labor force status of the aged. They differ primarily in the emphasis placed on the variables and direction of causality between social policy and participation. This section specifies in more detail the mechanisms through which development and social policy determine participation and discusses the differences in causality specified by the theories.

Economic Development and Participation of Aged Males

The industrialism theory argues that forces brought about by continued industrialization and economic growth have led to forced retirement of workers. Several components of industrialism are responsible for this change. First, developed nations have experienced a shift from a primarily agricultural, self-employed occupational structure to a wage and salary, bureaucratic one. Self-employed workers can choose how long they remain at their jobs; workers in larger bureaucratic organizations face formalized retirement rules based on chronological age (Slavick 1966). The shift to larger organizations will therefore reduce participation of older workers (Friedman and Orbach 1974).

Second, the expansion of the educational system in developed nations places older workers at a competitive disadvantage relative to younger workers in the labor market. Because of growth in educational opportunities, recent cohorts are likely to have completed more years of schooling than cohorts that completed schooling in earlier decades. Employers are likely to prefer younger workers with the latest skills and knowledge and to require the retirement of older workers (Clark, Kreps, and Spengler 1978).

Third, development has been accompanied by fertility changes that

result in a higher proportion of aged persons in the population. Given employers' preferences for younger workers, the increased number of older persons must compete for a shrinking pool of available jobs. With this increased competition, a lower proportion of older persons will be able to remain in the labor force (Cowgill 1974).

Finally, low demand for older workers (as reflected in unemployment levels) may reduce participation rates. Discouraged workers—unemployed persons who drop out of the labor force after unsuccessful job searches—are most common at the older ages (Rosenblum 1975; Clark et al. 1978). High unemployment may thus force older workers out of the labor force permanently. Although unemployment is not necessarily the direct result of economic development, it can be included with the industrialism variables because it is closely related to demand for older workers and involves involuntary retirement.

Given the effect of these components of economic growth on participation of aged males, an important implication follows. Since, according to the theory, the incidence of forced retirement is high, the effect of state social security expenditures and policies on participation will be limited. The retirement decisions of those who are forced to retire will not be influenced by public pension benefits and policies. Any association between expenditures by governments for pensions and the participation rate results from benefits responding to changes in participation rather than vice versa. Expenditures for public pensions (per retired person), instead of causing lower participation, may grow in order to limit social unrest (Isaac and Kelly 1981) and may respond to the pressures of interest groups who represent an already large retired population (Calhoun 1978). The argument is consistent with studies that show a substantial lag between the drop in participation caused by economic and industrial growth and the development of public pension benefits high enough to support retired persons in the United States (Clark et al. 1978), Britain (Quadagno 1982), and other developed nations (Palmore and Manton 1974). In sum, the theory predicts that the effect of public pension policies and expenditures will be limited compared with the effects of economic growth and that this effect may be overestimated owing to the simultaneous influence of participation on expenditures and policies.

Social Policy Determinants of Participation of Aged Males

The second explanation focuses on the primacy of variables related to state policies for pensions and social welfare.[3] Because persons desire

[3] The focus here is on public pension and social insurance expenditures of governments. Data on private expenditures for pensions, although not available in comparable form for even the most developed nations, are also likely to be related to participation. However,

leisure if they can afford it, pension and welfare expenditures of nations can attract persons out of the labor force. While economic development provides the needed surplus income for public pension systems, nations vary in the proportion of this surplus that goes to aged persons (Horlick 1970). Social policies can thus exert an influence on participation that is independent of economic development and thereby prevent the convergence of patterns of labor force participation of older males in advanced developed nations (Palmore 1975).

Four groups of government policies may be relevant to the participation rates of aged males. First, it is likely that the higher the government expenditures for pension and social insurance programs (per retired person), the lower the participation rate of aged males. Second, expenditures for nonretirement programs may also reduce participation. For instance, medical welfare payments, public housing, welfare programs, income supplements for food and energy, and various in-kind benefits may make the environment more conducive to retirement. While it is difficult to specify all the programs that indirectly support retired older persons, it is likely that the greater the support of the government for social welfare in general, the more retired persons will benefit and the lower the participation rate will be.

Third, certain provisions of pension programs may increase the number of persons retiring, even controlling for the level of pension and other government expenditures. For example, eligibility for pension benefits at or before age 65 may encourage early retirement; a retirement test that reduces pension benefits for those who continue working after the age of eligibility may encourage some to retire completely to obtain full benefits; or coverage of all workers or citizens by the pension system may provide more persons with the opportunity to receive retirement benefits.

Finally, the number of years since the inception of the pension program indicates the extent to which the program has been and is recognized as stable and reliable. The longer a program has been in existence, the greater influence it will have on the retirement decision and the participation rate of aged males (Wilensky 1975).

Changing Determinants over Time

The preceding two explanations may be consistent if a temporal scope is specified for each of them. Specifically, it may be that, during early time

---

the existence of private pensions is highly correlated with the existence of a mandatory retirement policy (Clark et al. 1978) and will be at least partly summarized by the economic growth variables described above. Furthermore, the level of private benefits relative to public benefits is low even in the United States, which traditionally has relied on the private sector more than most of the other developed nations (Schulz 1980).

periods, the effects of economic growth are the best predictors of labor force participation but in later time periods level of benefits and other characteristics of pension programs may have larger effects than economic development variables on participation. In short, this argument suggests the need to test for interaction among participation, its determinants, and time.

METHODS

Sample and Data

As Cutright (1967) argued, political and social differences among nations can affect social welfare expenditures and participation only in countries with levels of economic development high enough to afford a large institution of retirement. Theoretically, then, the appropriate population for our study is developed countries in recent decades. Given limitations in the availability of data, the sample chosen for study includes 18 nations for six time points—every five years between 1950 and 1975. The 18 countries were chosen because they are relatively developed economically and had all the necessary data.[4] The choice of the time points—1950, 1955, 1960, 1965, 1970, and 1975—was warranted by the publication of participation rates by age only at five-year intervals. Together, the combined cross section and time series provide a sample size of 108 cases that have relatively complete, reliable, and comparable measures of the necessary variables. Because some additional measures of pension system characteristics are available only from 1955 on, some of the analyses are for the smaller sample of 18 nations and five time points (90 cases). Similarly, an alternative measure of pension expenditures is available only for 1960 on, limiting these analyses to 72 cases.

Variable Measurement and Indicators

The dependent variable is the labor force participation rate of males aged 65 and over. Labor force participation as defined by the ILO is the percentage of the age-specific population economically active at the time of survey. The economically active population includes employed persons (paid workers and unpaid family workers) and unemployed persons ac-

[4] The countries included in the sample are Austria, Belgium, Canada, Denmark, Finland, France, West Germany, Ireland, Italy, Japan, Norway, Netherlands, Switzerland, Sweden, United Kingdom, United States, Australia, and New Zealand. Nations with planned, central economies are not included since data on gross domestic product and several other variables are not available.

357

tively seeking work. The figures come from ILO estimates that have been adjusted in all nations to accord with the standard definition.

Level of economic development is measured by the gross domestic product (GDP) per capita in constant (1970) dollars. Like the other monetary measures used in this study, the GDP will be translated into U.S. dollars by using data on the exchange rate.

Changes in the occupational structure are measured by the percentage of the labor force in agricultural occupations. This measure is closely related to the percentage of the work force who are self-employed and to the mean size of work organizations (Land and Pampel 1977); theory predicts that both of these will affect labor force participation. Thus, the size of the agricultural labor force is used as an indicator of general shifts to a bureaucratized labor force rather than as a measure only of changes in retirement in the agricultural sector. The measure has also been standardized and adjusted by the ILO and is comparable across the nations in this study.

Educational levels are measured by the percentage of males ages 20–24 enrolled in tertiary educational systems. The higher the enrollment, the more the young are advantaged relative to the old. This variable will also be measured as a change score that assumes that the greater the increase in enrollment, the greater the advantage of the young cohorts. Ideally, median education of older cohorts relative to the median education of younger cohorts should be used to measure competitive disadvantage of the aged in the labor market. Since such data are not available, we use a crude indicator of the size of the educational system in recent decades.

Age structure is measured by the ratio of males over age 65 to males age 25–64 (times 100). Unemployment is measured as the percentage of the labor force not currently at work. The measure includes all males and females in both the numerator and denominator; sex- and age-specific unemployment rates are not available for all the time points for all countries.

Social security benefits will be measured in two ways because one measure is available for all years from 1950 to 1975 while the other is available only for years since 1960. The first measure is public social insurance expenditures of the government. This is a broad grouping of expenditures, primarily of those for pensions, but also including sickness-maternity, employment injury, and unemployment benefits. The second measure includes only total public pension expenditures of a government and is available for a shorter time period. Because there are two measures covering different time periods, two separate analyses will be performed, one for the whole time period, using social insurance benefits, and one for the years since 1960, using pension benefits. Both measures are divided

by the number of males aged 65 and over not in the labor force in order to tap changes in average benefits.[5] Variations in the dollar metrics of the measures—such as the use of current dollars, constant dollars, or dollars relative to the domestic product per capita—will be examined to determine which best predicts participation.

The measures of other public pension systems provisions are based on classifications designed by the Social Security Administration in 1955. First, coverage of the male population is measured in four categories: (1) only employees with limited means; (2) wage and salary workers; (3) all employees; and (4) all citizens. Second, age of eligibility for benefits is measured as a dummy variable, with 1 indicating eligibility after 65 and 0 indicating eligibility for retirement at or before 65.[6] If the nation has a retirement test limiting benefits for those who continue to work after the retirement age, the retirement test dummy variable is coded 1. If a nation has a means test that provides special benefits to those with relatively low income, the means test variables are coded 1. Measures of the years since the first pension law was passed are also included as indicators of the institutional strength and stability of the system.

The size of non–social insurance transfer programs and the general welfare support of governments for their citizens will be measured by government revenues per capita. The greater the government revenues, the more involved governments are in transferring income among groups of the population and the greater role it takes in supporting low-income populations such as retired persons (Rubinson and Quinlan 1977). Like social insurance expenditures, government revenues can be measured in current dollars, constant dollars, and dollars relative to the standard of living (i.e., GDP).

Methods of Estimation

In using ordinary least-squares (OLS) estimation techniques with the pooled cross-section time-series design, there is a problem that the residuals may be serially correlated over time and cross-sectionally correlated across nations. Without a lagged dependent variable, OLS estimates are then inefficient; with a lagged dependent variable, they are biased. Moreover, correlated residuals may be seen as the result of specification error

---

[5] Data on the ideal denominator, the number of social insurance recipients, are not available. Other denominators, such as "all persons over age 65" or "all males over age 65," were tried but did not have effects substantially different from those reported here. Finally, since more sophisticated measures of benefits such as the average replacement rate of pensions for wages were not available for most of the nations and time points, we therefore rely on expenditures divided by "potential male recipients."

[6] Since the focus is on participation of persons aged 65 and over, variations in early retirement options are not important in differentiating pension program characteristics.

or neglected variables (Hannan and Young 1977). For these reasons, it may be necessary to adjust for correlated residuals by using modified generalized least squares (MGLS).

We will use two forms of MGLS which vary in the assumptions made about the error term. The first method, termed the random error components model, assumes the error term to be made up of three components: one that is constant over time but varies across nations and summarizes the autocorrelation of residuals within a nation; one that is constant across countries but varies over time and summarizes the cross-sectional correlation of the residuals; and one that is random. Once estimates of the error components are obtained from variances and covariances of the OLS residuals, the usual generalized least-squares (GLS) procedures are used to adjust the estimates (Fuller and Battese 1974). The major implication of the model is that it assumes that the correlation between residuals within one nation is the same between all time points (e.g., the residual correlation between 1960 and 1965 is the same as between 1960 and 1975).

The second model assumes an autoregressive process wherein the residuals are a function of values at the previous time point and an autoregression coefficient unique to each country. The residuals are also a function of cross-sectional correlation and random error. Unlike the previous model, the autoregressive model assumes that the correlation between residuals is a function of the previous year's values. Ordinary least-squares estimates are used to estimate the autocorrelation coefficient and the cross-sectional correlation, which are then used with generalized least squares to adjust the estimates (Kmenta 1971).[7]

CHANGES IN PARTICIPATION RATES

Before discussing estimation of the causal model, we describe the extent of change in labor force participation and its relationship to economic growth. The mean participation rate declines from 37.9 in 1950 to 20.6

[7] Two related problems in estimation involve heteroscedasticity and use of ratio variables. Without the use of ratio variables, countries with large populations are likely to have large error variance. The use of ratio variables standardizes for population size and size of the economy but may create spuriousness between ratios that have common components. In the following analysis, common components exist only in the number of males age 65 and over in the numerator of the old-age dependency ratio and the denominator of the participation rate. Yet, both these variables can be best interpreted theoretically as ratios; i.e., our interest is in the size of the aged population relative to the non-aged population as a measure of competition, and our interest is the probability or rate of participation rather than in the raw numbers. Given this goal, the use of ratio variables is most appropriate. Moreover, the use of common components is not likely to exaggerate the relationship since the correlation between number of males aged 65 and over in the labor force and the number of aged males is .95. Use of ratio variables with these components reduces the correlation substantially. Thus, the strategy to be used is based on ratio variables.

360

in 1975 {a grand mean of 28.1; table 1). The GDP per capita in thousands of U.S. dollars increases from 1.54 to 3.76 (a grand mean of 2.34 thousand dollars per person). The linear relationship between the two is shown by a correlation of − .37 (table 1). The multiple correlation is increased to .409 with the inclusion of a quadratic term. While the additional variance explained is small, the unstandardized coefficient for the squared term is positive and significant, which indicates that the decline in participation levels off with continued development.

While the participation rate declines logistically with economic development, intercountry differences in participation are not lower for later time periods or for higher levels of economic development. Over time, the standard deviation of the participation rate has remained constant, and the coefficient of variation (the standard deviation divided by the mean) increases. When the data are grouped into six categories on the basis of economic development, the mean participation rate is lowest for the high-GDP group, while the coefficient of variation is highest. Overall, with continued economic development, there is considerable fluctuation in the coefficient of variation and little evidence, of convergence in participation rates.

To summarize, level of economic product shows a modest effect on the participation of older men—an effect limited here by the truncated sample of developed nations. As a result, much of the variance in participation remains to be explained by other measures of economic development and social welfare expenditures and policies.

DETERMINANTS OF PARTICIPATION

The means, standard deviations, and correlations with the participation rate of older males for the independent variables are listed in table 1. The variables can be divided into three groups: (1) the components of economic growth; (2) monetary measures of total government revenues and of expenditures for social insurance and pension programs; and (3) other nonexpenditure characteristics of the public pension system. A number of variables in each group are related to participation—primarily the size of the agricultural labor force, social insurance expenditures, and age of the public pension system. Because these variables are intercorrelated, their net effects need to be examined. An advantage of the pooled cross-section time-series design and the truncated sample, however, is that the sizes of most of the correlations are small compared with those existing in time-series or cross-sectional aggregate data. All but five of the more than 100 correlations among the independent variables are below .5, indicating that multicollinearity will not be a serious problem (none of the collinear variables are included in the same equation).

The analysis begins with consideration of the determinants of partic-

361
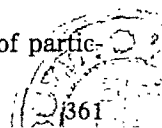
TABLE 1

MEANS, STANDARD DEVIATIONS, CORRELATIONS WITH LABOR FORCE PARTICIPATION
OF MALES AGE 65 AND OVER, and DEFINITIONS OF VARIABLES, 1950–75

Variables	Mean	SD	r	Definition Units
Labor force participation (male 65 +)	28.1	11.7	1.00	Percentage of males age 65 and over in the labor force
Percentage agricultural	17.9	10.9	.69	Percentage of male labor force in agricultural occupations
Enrollment tertiary	6.4	4.6	−.37	Percentage of males aged 20–24 enrolled in tertiary schools
Old-age ratio	18.6	3.3	−.32	Ratio of males age 65 + to males 25–65 times 100
Unemployed	2.3	2.1	−.03	Percentage of persons in the labor force unemployed
GDP (US$)	2.34	1.15	−.37	GDP in thousands of 1970 US$ per capita
Social insurance expenditures (US$). . .	4.30	3.24	−.50	Thousands of 1970 US$ of social insurance expenditures per male age 65 + not in the labor force
Social insurance expenditures/GDP . . .	1.79	₂79	−.46	Ratio of social insurance expenditures per retired male to GDP per capita
Pension expenditures (US$)*	2.95	1.82	−.45	Thousands of 1970 US$ pension expenditures per male age 65 + not in the labor force.
Pension expenditures/ GDP*	1.06	.40	−.35	Ratio of pension expenditures per retired male to GDP per capita
Government revenue (US$)	.47	.32	−.42	Thousands of 1970 US$ of government revenues per capita
Government revenue/ GDP.	.19	.07	−.24	Ratio of government revenues to GDP
First law†	48.7	18.0	−.45	Years since first public pension law
Coverage†	3.11	1.11	−.04	Rank order scale of coverage of population by public pension system (see text)
Age retirement†	.24	.43	.33	Eligibility for benefits after age 65 = 1; age 65 or before = 0
Retirement test†	.48	.50	−.27	Retirement test = 1; none = 0
Means test†	.67	.47	.31	Means test = 1; none = 0

* Covers only years from 1960 to 1975.

† Covers only years from 1955 to 1975.

ipation for the full time period, 1950–75. This necessitates (1) the exclusion of the nonexpenditure public pension policy characteristics which are available only from 1955 on, and (2) the use of the social insurance expenditures variable instead of the pension expenditures variable which is available only from 1960. Analysis of the additional variables for the shorter time span will follow.

Table 2 presents models of participation, using the specified variables under several different estimation techniques. The equations include several measures related to economic development: the percentage of the male labor force in agricultural occupations, the percentage of young males enrolled in tertiary schools, the percentage unemployed, and the male old-age dependency ratio. The direct measure of economic growth, gross domestic product per capita in constant U.S. dollars, is not included in the equation. Our rationale for this is that the other variables are predicted to mediate the effect of economic growth. Indeed, gross domestic product correlates −.57 with the percentage in agricultural labor force and .76 with tertiary enrollment. Inclusion of all three variables leads to unstable estimates and insignificant effects of the gross domestic product. Hence, we use the more meaningful theoretical components of economic growth instead of the direct measure. The equations also include two policy-manipulable variables, social insurance expenditures and government revenues. Some equations use both measures in constant U.S. dollars per capita, while others use both measures relative to the gross domestic product. The different equations will provide a comparison of the effects on participation of absolute and relative benefit level measures.[8]

The OLS estimates in table 2 show that the percentage of males in agricultural occupations has the largest effect on participation of aged males. A decline of 1% in agricultural employment reduces participation by about .677%. None of the other variables resulting from economic growth affects participation significantly. The second strongest effect thus comes from the social insurance expenditure measure. Comparison of the first and second equations shows that the measure for social insurance expenditures relative to the gross domestic product has a larger standardized coefficient and explains more variance than the absolute social insurance expenditure measure. As an inducement to retirement, the level of social insurance benefits needs to keep pace with the standard of living of the population as well as with inflation to have its strongest effect on participation. For both the absolute and ratio measures, however, the effect is linear and does not show a threshold. If there is a level of benefits

---

[8] The effect of expenditures in current dollars is smaller than that of either expenditures in constant dollars or expenditures relative to the gross national product and is not included in the results.

TABLE 2

UNSTANDARDIZED AND STANDARDIZED COEFFICIENTS FOR OLS AND GLS ESTIMATES
OF MODELS OF PARTICIPATION OF AGED MALES, 1950–75

	Ordinary Least Squares		Generalized Least Squares	
	(1)	(2)	Error Component	Autoregression
Constant	23.191	27.334	40.920	31.992
Percentage agricultural	.677*	.658*	.609*	.540*
	.628	.610	.565	.500
Old-age ratio	−.254	.223	−.856*	−.303***
	−.073	.064	−.245	−.087
Enrollment tertiary	.389	.153	.167	−.106
	.153	.060	.065	−.042
Unemployed	−.342	−.355	−.201	−.175
	−.060	−.062	−.035	−.031
Social insurance expenditures	−.662**	...	...	...
	−.183			
Government revenue	−2.929	...	...	...
	−.079			
Social insurance expenditures/GDP	...	−5.743*	−3.232*	−2.545*
		−.384	−.217	−.171
Government Revenue/GDP	...	−27.054*	−14.065	−8.200
		−.206	−.107	−.062
R^2	.5303	.6117	...	...
Durbin-Watson	.429	.528	...	...
df	101	101	101	101

NOTE.—Standardized coefficients are presented below unstandardized coefficients.

* $P < .01$.

** $.01 < P < .05$.

below which few persons will retire voluntarily, it does not show up in this sample.

The effect of government revenues also is stronger when measured in relative terms. The higher the level of revenues, the lower the participation rate. This suggests that the larger the influence of the state in providing for the social welfare of the population (net of social insurance benefits), the better are the opportunities for aged males to leave the labor force.

In the OLS equations, the Durbin-Watson statistic does not allow acceptance of the null hypothesis of no autocorrelation. Although the OLS estimates are unbiased, they are inefficient, and the relationships need to be examined once the adjustments for autocorrelated residuals have been made. Error components and autoregressive MGLS estimates are presented in table 2.

There are a number of similarities in the OLS and MGLS estimates. The effects of agricultural labor force and of social insurance expenditures remain significant, and the relative expenditures variables continue to have stronger effects than the absolute measures. However, there are also some important differences. The male old-age dependency ratio, not significant in the OLS estimates, is significant in the MGLS estimates. As predicted, the higher the ratio of older males to younger adult males, the lower the participation rate of older males. The government revenues measure, significant in the OLS models, is no longer significant in the MGLS models. Its significance in the OLS estimates may result from unmeasured variables related to government revenues, the effects of which are removed once we make adjustments for serial correlations. We will attempt to identify some of these neglected variables in subsequent sections.

A variation in the models in table 2 involves the use of relative social expenditures lagged one time period (i.e., five years). Since the time lag has implications for the causal relationship between social insurance benefits and participation, we explore this relationship in more detail in table 3. In this table, trimmed versions of the participation equations from which consistently insignificant tertiary enrollment and unemployment variables have been deleted are presented with and without a lagged relative social insurance expenditures measure. The use of a lagged variable reduces the time points by one and the degrees of freedom by 18, so this analysis is based on data from 1955 to 1975. The first three equations, which do not include the lagged variable, replicate the findings from table 2. Yet, when both the lagged and unlagged relative social insurance expenditures variables are put into the equations, the effect of the lagged measure is stronger. In fact, the unlagged measure is not significant in the OLS or error component estimates when the lagged measure is con-

TABLE 3

UNSTANDARDIZED AND STANDARDIZED COEFFICIENTS FOR OLS AND GLS ESTIMATES OF LAGGED MODELS OF PARTICIPATION OF AGED MALES 1955–75

	UNLAGGED SOCIAL INSURANCE/GDP			LAGGED SOCIAL INSURANCE			LAGGED SOCIAL INSURANCE GDP AND LAGGED MLFP65 +
	OLS	Error Components	Autoregression	OLS	Error Components	Autoregression	OLS
Constant	29.060	39.759	28.203	26.854	39.892	40.394	1.16
Social insurance/GDP	-5.861*	-2.442*	-2.773*	-1.939	-1.130	-1.605**	...
	-.422	-.176	-.200	-.140	-.082	-.116	...
Social insurance/GDP t-1	...	...	...	-5.313*	-2.557*	-3.711*	-1.30**
				-.364	-.175	-.254	-.089
Percentage agricultural	.596*	.557*	.670*	.579*	.515*	.419*	-.098***
	.536	.501	.602	.521	.464	.377	-.088
Old-age ratio	.230	-.829*	-.380*	.474	-.655*	-.536*	.169
	.072	-.258	-.118	.147	-.204	-.167	.053
Government revenue/GDP	-30.414*	-11.793	-.389	-32.871*	-16.383	-13.819	-10.808**
	-.251	-.097	-.003	-.271	-.135	-.114	-.089
MLFP65 + t-1	...	...	...	...	...	...	.934*
							.974
R^2	.5698	...	...	.5992	...	...	.9258
Durbin-Watson	.625	...	...	.570	...	...	1.169
df	85	85	85	84	84	84	84

NOTE.—Standardized coefficients are presented below unstandardized coefficients.

* $P < .01$.

** $.01 < P < .05$.

trolled for and can be deleted from the model without reducing the explained variance significantly. This result shows some gap between the time persons use information on public retirement to make the retirement decision and the time they actually retire. Although the five-year lag is not necessarily the most accurate (one, two, three, or four years might have worked better, had yearly data been available), the results do provide some evidence that the effect of expenditures precedes participation.

In summary, the percentage in the agricultural labor force has the strongest effect on participation, followed by the effects of social insurance expenditures and the male old-age dependency ratio. The substantially different results for the OLS and the MGLS estimates suggest the need to adjust for serial correlation. Further, the models are primarily additive and instantaneous. The models in table 3 explain as much or more variance than: (1) multiplicative models where all variables are transformed into their natural logarithms; (2) curvilinear models where the independent variables are transformed into their natural logarithms; (3) change models where the independent variables are measured as difference scores (from t to $t - 5$); and (4) lagged models where the independent variables affect participation for the next five years.[9]

A final variation on the previous results is to include a lagged endogenous variable as a predictor. Although the previous results showed that the effect of expenditures precedes participation, this does not prove there is no simultaneity bias in the relationship. It may be that the relationship between expenditures lagged and participation is spuriously attributed to the common cause of participation lagged. To test this possibility, participation lagged can be used as an independent variable, and spuriousness between participation and expenditures lagged can be removed. In addition, the use of the lagged participation variable provides a model in which the independent variables predict change in the dependent variable (i.e., level of participation with controls for the previous year's level of participation). This can provide a further test of the ability of the independent variables to explain both the levels and the change in participation over time.

The OLS estimates for equations with the lagged dependent variable are presented in table 3. The effect of participation lagged is strong: the standardized coefficient is close to one, the variance explained is increased to above 90%, and the effect of all the other variables is reduced. However, the effect of social insurance expenditures remains significant. This

---

[9] Another issue in the construction of this model concerns the constancy of the causal processes over the different panels. It is recommended that models be estimated for the separate panels to test for homogeneity of the causal processes and the coefficients (Hannan and Young 1977). Tests for interaction over time will be considered in the next section.

provides further evidence of the causal effect of the variable.[10] However, the model has disadvantages compared with others in table 3. For instance, the strong autocorrelation of participation leaves little variation to be explained by the theoretically more important variables. Furthermore, the OLS estimates are inconsistent as well as inefficient, given the Durbin-Watson statistic of 1.16 and the presence of a lagged dependent variable.[11] Because of these problems, and the similarity of the results with and without the lagged dependent variable, the models that follow do not include participation lagged.[12]

EFFECTS OF OTHER PENSION POLICY CHARACTERISTICS

So far the results have shown the dominant determinant of participation of aged males to be the percentage of the labor force in agricultural occupations. While social insurance expenditures have a nonspurious, negative effect on participation, it is typically less than half the size of the agricultural effect. It is possible that the effects of the policy variables would be stronger if: (1) measures of other nonexpenditure provisions of pension systems were included in the model; (2) pension expenditures rather than social insurance expenditures were used; and (3) interaction

[10] A second way to look at this issue is to estimate a simultaneous feedback model in which participation is allowed to determine expenditures at the same time expenditures determine participation. Such a model would require the choice of instrumental variables as well as specification of a model for social insurance expenditures to see if these are determined by participation. Development of such a model requires a paper in itself, and we cannot present a full and complete consideration of the reasoning and results here. Nonetheless, we have experimented with two-stage least-squares estimates (unadjusted and adjusted for autocorrelation), and information on the results can be obtained from the senior author. The results show that simultaneity bias does not appear to be responsible for the effects of social insurance expenditures on participation. However, these results are clearly preliminary. Much more attention needs to be given to development of instrumental variables and the determinants of expenditures before strong conclusions can be reached.

[11] The Durbin-Watson statistic is biased toward the acceptance of the null hypothesis of no serial correlation in the presence of lagged dependent variable. The actual serial correlation is likely to be stronger than that estimated by the Durbin-Watson statistic. Thus, at the least, we can be more certain that the estimates are biased and that autocorrelation exists. Although it was not designed for pooled cross-sectional data, calculation of Durbin's h also shows significant serial correlation.

[12] Adjustment for autocorrelation with MGLS is difficult in equations with a lagged dependent variable because of inaccurate estimation of residual terms when the model is biased. Moreover, Hannan and Young (1977) argue that, in small samples with a lagged dependent variable and autocorrelation, OLS may outperform MGLS in terms of bias and mean squared error. A further alternative is to use instrumental variables to remove the correlation between lagged dependent variable and the error term. Typically, the exogenous variables lagged are used as the instruments. Although Hannan and Young (1977) recommend against these models since lagging the exogenous variables eliminates one wave of cases and is less efficient than modified MGLS, we did experiment with the procedure. The predicted lagged dependent variable was insignificant and negative in the OLS and MGLS estimates and therefore did not suggest the need to use a lagged dependent variable in other models.

368

models, which allow the effects of pension or social insurance expenditures to increase over time, were estimated. This section examines each of these hypotheses in turn.[13]

First, we estimated models of participation of aged males which include the five measures of provisions discussed earlier: age of eligibility; existence of a retirement test; existence of a means test; coverage of the population; and years the program has been in existence. The data limitations confine the time span to the years since 1955. Although the variables are correlated with participation (table 1), the MGLS estimates show that none of the pension provisions affects participation significantly at the .01 level (existence of a means test is significant only at .05 in one of the MGLS equations). Moreover, the effects of percentage in agriculture remain strong and significant when pension system characteristics are controlled for. The addition of other pension system characteristics, then, adds little to the model or to the explanation of participation.

Second, the effects of social insurance benefits may be small relative to size of the agricultural labor force because a poor measure of benefits is used. A superior measure would be actual expenditures for pensions, a measure available only for the years from 1960 to 1975. The models in table 3 were estimated for this period, using pension expenditures as a ratio to the gross domestic product in one set of equations and social insurance expenditures as a ratio to the gross domestic product in another set. In all but one of the equations, the standardized coefficient for the social expenditures ratio is larger than the standardized coefficient for the pension expenditures ratio. Thus, for this time period and sample nations, the social insurance expenditures ratio predicts participation at least as well as the pension expenditures ratio. The previous results found for social insurance expenditures over a longer time span would likely hold, even with a measure of pensions.

Third, even if the effects of social insurance expenditures are smaller than the effects of percentage in agriculture, over time the trend may be toward larger policy effects and smaller industrialism effects. To test this hypothesis, multiplicative interaction terms (time multiplied by each of the variables in table 3) were included with the additive terms. While OLS results do not show that the interactive terms increase the explained variance significantly, the MGLS estimates are slightly more supportive of the interaction hypothesis: over time the effects of percentage in agriculture decline. Yet, the effects of social insurance expenditures do not increase, and the effects of percentage in agriculture remain as strong as those of social insurance expenditures for even the most recent periods.

[13] The tables on which the following discussion is based may be obtained from the first author.

It may be that percentage in agriculture has reached a floor in recent time periods in several nations so that further changes can have little effect on participation. Despite this change, however, the effects of social policy variables are not dominant.[14]

SUMMARY AND DISCUSSION

An initial distinction was made between the effects on participation of aged males of long-term trends related to industrialization and of short-term changes in public pension programs designed to make retirement more attractive. Of the variables tested with several models, estimation techniques, and time periods, the percentage of the male labor force employed in agricultural occupations—which summarizes changes in bureaucratization, loss of choice in retirement age, and the needs of industry for younger workers—had the largest, most consistent effects. The male old-age dependency ratio had smaller but significant effects. Both these variables show the importance of changes in economic demand for workers in explaining the decline in participation of aged males in developing nations.

Among the policy variables, the major effect on participation comes from levels of social insurance expenditures relative to the gross domestic product. The effect was found to be causally prior to participation and at least partially responsible for continued divergence in participation rates for this sample of nations over the past 30 years. Other policy variables, however, such as government revenues and pension system provisions, did not significantly affect participation of aged males.

Theoretically, the results support the logic of the industrialism argument that emphasizes the low demand of organizations in developed economies for older workers. Theories of modernization and aging have been criticized because they have an oversimplified focus on a single factor and cannot account for deviations from their predictions (Palmore 1975; Laslett 1976; Achenbaum and Stearns 1978; Dowd 1981). Yet, this paper shows that one variable—size of the agricultural labor force—has a crucial impact on participation and that, at the societal level, technology strongly predicts the age structure of the labor force. While single-factor theories, however parsimonious they may be, cannot explain variation in national characteristics such as labor force participation completely, industrialism or technological determinism theories of retirement work better than recent literature suggests.

The qualification to the theory provided by the results in this paper is that differences in social insurance expenditures do account for some

[14] Other interactions of the variables with GDP and years since the pension law has been in existence were insignificant.

variation in participation and may maintain divergencies in participation rates. While all nations experience a decline in participation with economic growth, differences in social insurance expenditures maintain the relative positions of nations in the face of increasingly similar industrial structures. Perhaps this suggests a weaker version of the industrialism theory, in which all nations eventually arrive at low levels of participation but some nations rely more than others on supply or social policy means. The needs of the industrial economy for a younger, more educated labor force may be met primarily through the private sector response of mandatory retirement; yet, some variation in participation remains because of national differences in the public sector response of higher retirement benefits.

Another implication of the results is that the ability of governments to reverse the trends in participation appears to be limited. Although drastic cuts in benefits would raise participation, more modest changes in benefits, retirement tests, coverage, or age of eligibility would have small effects. Without concomitant changes in the labor demands of the economy, participation of aged males in developed nations is likely to remain low.

Methodologically, use of the pooled cross-section time-series approach to reach these conclusions improves upon the previous studies of single cross sections or single time series. The need to examine a larger sample of countries with better measures of some of the variables remains. Further research needs to examine the emergence of systems of retirement across developed and developing countries. Yet, the advantages of a design that provides variation across time and nations warrant the further use of the pooled cross-section, time-series design.

REFERENCES

Achenbaum, W. A., and P. N. Stearns. 1978. "Old Age and Modernization." *Gerontologist* 18:307–12.
Bowen, William G., and T. Aldrich Finegan. 1969. *The Economics of Labor Force Participation*. Princeton, N.J.: Princeton University Press.
Burgess, E. W. 1960. "Aging in Western Culture." Pp. 3–28 in *Aging in Western Societies*, edited by Ernest W. Burgess. Chicago: University of Chicago Press.
Calhoun, Richard B. 1978. *In Search of the New Old: Reaffirming Old Age in America.* New York: Elsevier North-Holland.
Clark, R., J. Kreps, and J. Spengler. 1978. "Economics of Aging: A Survey." *Journal of Economic Literature* 16:919–62.
Cohn, R. M. 1982. "Economic Development and Status Change of the Aged." *American Journal of Sociology* 87:1150–61.
Cowgill, D. O. 1974. "Aging and Modernization: A Revision of the Theory." Pp. 123–46 in *Late Life, Communities and Environmental Policy,* edited by Jaber F. Gubrium. Springfield, Ill.: Thomas.
Cowgill, D. O., and Lowell D. Holmes. 1972. *Aging and Modernization.* New York: Appleton-Century-Crofts.

Cutright, Phillips. 1967. "Political Structure, Economic Development and National Security Programs." *American Journal of Sociology* 70:537–48.

Dowd, J. J. 1981. "Industrialization and the Decline of the Aged." *Sociological Focus* 14:255–70.

Durand, John D. 1975. *The Labor Force in Economic Development*. Princeton, N.J.: Princeton University Press.

Fisher, P. 1978. "The Social Security Crisis: An International Dilemma." *Aging and Work* 1:1–14.

Form, William H. 1979. "Comparative Industrial Sociology and the Convergence Hypothesis." *Annual Review of Sociology* 5:1–25.

Friedman, E. A., and H. L. Orbach. 1974. "Adjustment to Retirement." Pp. 604–45 in *American Handbook of Psychiatry*, edited by Silvano Arieti. New York: Basic.

Fuller, W. A., and G. E. Battese. 1974. "Estimation of Linear Models with Crossed-Error Structure." *Journal of Econometrics* 2:67–78.

Graebner, William. 1980. *A History of Retirement: The Meaning and Function of an American Institution, 1885–1978*. New Haven, Conn.: Yale University Press.

Griffin, C. J., J. A. Devine, and M. Wallace. 1981. "Accumulation, Legitimation, and Politics: The Growth of Welfare Expenditures in the United States since the Second World War." Paper presented at the annual meetings of the American Sociological Association, Toronto.

Hannan, M. T., and A. A. Young. 1977. "Estimation in Panel Models: Results on Pooling Cross-Sections and Time Series." Pp. 52–83 in *Sociological Methodology 1977*, edited by D. Heise. San Francisco: Jossey-Bass.

Horlick, M. 1970. "The Earnings Replacement Rate of Old-Age Benefits: An International Comparison." *Social Security Bulletin* 33:3–16.

Isaac, L., and W. R. Kelly. 1981. "Racial Insurgency, the State, and Welfare Expansion: Local and National Evidence from the Postwar United States." *American Journal of Sociology* 86:1348–86.

Kerr, Clark, John T. Dunlop, Frederick H. Harbison, and Charles A. Meyers. 1960. *Industrialism and Industrial Man*. Cambridge, Mass.: Harvard University Press.

Kmenta, Jan. 1971. *Elements of Econometrics*. New York: Macmillan.

Land, K. C., and F. C. Pampel. 1977. "Indicators and Models of Changes in the American Occupational System, 1947–73: Some Preliminary Analyses." *Social Indicators Research* 4:1–23.

Laslett, P. 1976. "Societal Development and Aging." Pp. 87–116 in *Handbook of Aging and the Social Sciences*, edited by Robert H. Binstock and Ethel Shanas. New York: Van Nostrand Reinhold.

Palmore, Erdman. 1975. *The Honorable Elders: A Cross-cultural Analysis of Aging in Japan*. Durham, N.C.: Duke University Press.

Palmore, Erdman, and K. Manton. 1974. "Modernization and Status of the Aged: International Correlations." *Journal of Gerontology* 29:205–10.

Pampel, Fred C. 1981. *Social Change and the Aged: Recent Trends in the United States*. Lexington, Mass.: Lexington.

Quadagno, J. S. 1982. *Aging in Early Industrial Society: Work, Family, and Social Policy in Nineteenth-Century England*. New York: Academic Press.

Rosenblum, M. 1975. "The Last Push: From Discouraged Worker to Involuntary Retirement." *Industrial Gerontology* 2:14–22.

Rubinson, R., and D. Quinlan. 1977. "Democracy and Social Inequality: A Reanalysis." *American Sociological Review* 42:611–23.

Schulz, James H. 1980. *The Economics of Aging*. Belmont, Calif.: Wadsworth.

Sheppard, H. L. 1976. "Work and Retirement." Pp. 286–309 in *Handbook of Aging and the Social Sciences*, edited by Robert H. Binstock and Ethel Shanas. New York: Van Nostrand Reinhold.

Slavick, Fred. 1966. *Compulsory and Flexible Retirement in the American Economy*. Ithaca, N.Y.: Cornell University Press.

Wilensky, Harold L. 1975. *The Welfare State and Equality*. Berkeley and Los Angeles: University of California Press.

The Black-and-White Model of Attitude Stability: A Latent Class Examination of Opinion and Nonopinion in the American Public[1]

Marylee C. Taylor
Pennsylvania State University

Latent class analysis is used to examine Philip Converse's propositions about political opinion and nonopinion in the American public. Data were taken from the 1956, 1958, 1960 and the 1972, 1974, 1976 American voter panel studies conducted by the University of Michigan Institute for Social Research. Results support Converse's black-and-white model of attitude stability which posits the existence of some stable opinion holders and a group of "nonopinion" respondents whose positions vary unpredictably over time. Rates of opinion holding in the population were estimated to range from a low of 19% for the utilities and housing item, which was Converse's focus, to an atypical high of 59% for the busing item in the second panel. Models allowing for true score change estimated that the number of true score changers was small. The nonopinion respondents were apparently responsive enough to item content that their aggregated responses reflected the climate of opinion among those whose attitudes were better crystallized. As Converse had concluded, there was relatively little overlap of "issue publics"—those who held opinions on one issue were not necessarily opinion holders on other questions. There were not consistent differences in education or political involvement between opinion holders and nonopinion respondents. Relationships of education and political involvement to opinion holding suggest that respondents hold stable opinions on those issues that have greatest personal relevance for them.

INTRODUCTION

The nature of political thinking among the citizenry is a matter of wide-spread interest to political analysts as well as to social scientists. Philip Converse has outlined an intricate set of propositions concerning the mix of true opinion and "nonopinion" in the public, the strength of relationships among political attitudes, and the prevalence of ideological thinking. The most central Converse propositions are not amenable to testing by conventional data analytic techniques, but they are uniquely suited to latent class analysis. Application of the latent class methodology to Converse's opinion and nonopinion hypotheses is reported here; a subsequent paper will treat the topics of ideological thinking and attitude interrelations.

The Converse Analysis

The 1964 Converse discussion of American political attitudes provoked a controversy now in its eighteenth year. Two themes from that early paper have been spotlighted in the continuing debates: first, that the organization of political beliefs and attitudes in the minds of the mass public differs markedly from the organization which is characteristic among political elites; second, that, despite obliging responses to political pollsters, large numbers of the American public actually have no meaningful opinions on many of the topics represented in political attitude research.

It is not surprising that there has been critical reaction to assertions widely interpreted as a challenge to political attitude research, and to democratic theory as well. But the vigorous discussion of these provocative ideas has often eclipsed details of the complex analysis which Converse offered in 1964 and has elaborated since (Converse 1970, 1974, 1975, 1980; Converse and Markus 1979).

The primary empirical base for Converse's claims was studies of the American electorate conducted at regular intervals since the 1940s by the University of Michigan Institute for Social Research. The three-wave 1956, 1958, 1960 panel study figured prominently in the initial (1964) paper, while the most recent article (Converse and Markus 1979) reconsiders earlier propositions in the light of 1972, 1974, 1976 panel results.

Building on conclusions drawn from American voter study open-ended interviews (Campbell et al. 1960), Converse's 1964 paper suggests, first, that only a small proportion of the American public holds political opinions that are embedded in an overarching ideological framework. In that

paper and in a later one (1975), Converse speculates that the use of such an abstract organizing framework—usually expressed in terms of a liberal-conservative dimension when revealed at all—is a function not only of political involvement but also of the cognitive development that advanced education encourages. By implication, ideological thinking is associated with socioeconomic position and is relatively immutable over the adult life span, although Converse concedes that intense political involvement can produce ideological development even among the less educated (1975, p. 103).

Second, Converse claims that constraint among belief elements, that is, the interrelationship of political attitudes, is more prevalent than ideological thinking (and more democratically distributed), is more changeable over the individual lifetime, and is the product of a broader range of influences. Abstract ideological frameworks are only one source of constraint. Others include the coincidence of self-interest across multiple issues and the clustering of messages in the political diffusion process (Converse 1964, pp. 211–13). Primary predictors of attitudinal constraint are political motivation, interest, and involvement—presumably because these increase attention to opinion leaders' "packages" of positions as well as recognition of one's own interests vis-à-vis government policy. Some sets of political opinions are most tightly interrelated among highly educated and politically sophisticated ideologues who uniquely discern the relevance of these opinions to overarching political principles. This was the point Converse (1964) made in comparing attitude intercorrelations among congressional candidates with those among the public. But other sets of political attitudes, especially those sharing a concrete and specific social referent (e.g., racial minorities), may show greater constraint among the mass public than among political elites (Converse 1964, pp. 234–38).

The stability of individuals' political survey responses has been a third major focus of Converse's writing, less because of his interest in actual opinion shifts than because instability of *measured* opinion may indicate that *actual* opinions do not exist. Converse's most detailed discussions of the stability issue (1964, 1970, 1974) have centered on a 1956, 1958, 1960 voter study opinion item that asks whether government or private businessmen should be responsible for things like housing and utilities. Relative to its companion questions, this item is exceptional in several respects: it is remote from the immediate concerns of the average citizen; on its face, it seems closely tied to presumably stable fundamental ideological positions; it receives an unusually high rate of nonresponse; and the responses to this item at two consecutive waves of the panel are no more strongly related than responses at the first and third waves. This question inspired Converse's "black-and-white" model, which portrays respon-

dents as falling into two distinct classes: some hold a stable, presumably ideologically derived opinion and respond entirely consistently over time; others have no opinion and respond at random. Application of the black-and-white model to test-retest correlations on the housing and utilities item yielded an estimate that less than 20% of the American public actually held an opinion on this issue, although most non-opinion-holding survey respondents offered an opinion when asked (Converse 1970).

Converse acknowledges fully that this item is an extreme case. On most issues, he suggests, fully consistent and nonopinion respondents are joined by a "third force," a group experiencing true attitude change over the period of the three-wave study (Converse 1964, 1970). But Converse does propose that a modified version of the black-and-white model is widely applicable, that opinion data on many political issues are contaminated by a substantial proportion of nonopinion response—a disquieting possibility for those accustomed to taking these data at face value.

Who are the relatively stable opinion holders and who are the random nonopinion responders? If a single predictor must be used, partisan political activism, not education, is the best candidate (Converse 1975). But this general answer is not very satisfactory. No one group holds well-crystallized attitudes across all issues; nor can issues be ordered to form a Guttman scale, with each succeeding item expanding the core of stable opinion holders. Rather, the population contains a constellation of "issue publics," with overlap that is greater than chance but only moderate in absolute terms. The attitude crystallization reflected in stable responses is a function of the cognitive and motivational centrality of the particular issue for the individual, and refined predictions about attitudinal stability necessitate knowledge of individuals' interest in particular issues (Converse 1964).

Finally, how does response stability relate to constraint, to intercorrelations of responses across items? Stability and constraint apparently coexist—interitem relationships were elevated among respondents showing consistency over time on those items (Converse 1964). Converse (1975) takes this relationship to reflect the impact of constraint on stability, a presumption supported by cognitive social psychology evidence that attitudes linked to other cognitive elements are particularly resistant to change (see, e.g., Abelson et al. 1968). However, to the extent that observed stability reveals the existence of opinion rather than nonopinion, it is reasonable also to consider stability as a *condition* of constraint; lawful relationships should not be expected among nonattitudes.

The Continuing Debate

Debate over the Converse propositions has taken several forms:
1. The notion that ideological thinking is conditioned by the terms of

current political debate (Field and Anderson 1969) undergirds arguments that Converse's original paper referred to a period when ideological content in the public political arena was unusually low; the events of the 1960s are claimed to have increased attitudinal structure and/or ideological content in the American public's thinking (Bennett 1973; Margolis 1977; Nie and Anderson 1974; Pierce 1970; Pomper 1972; Stimson 1975).

2. The attitudinal stability or organization within subsamples, typically defined on the basis of education and/or political involvement as representing elites and the mass public, has been examined by correlational analyses (e.g., Bennett 1973; Bishop 1976; Bishop, Hamilton, and McConahay 1980; Stimson 1975), factor analyses (e.g., Luttbeg 1968; Kritzer 1978; Stimson 1975), and the Indscale technique (Marcus, Tabb, and Sullivan 1974; Stimson 1975). Some (e.g., Bishop et al. 1980) have attempted to circumvent the allegedly overrestrictive assumption (Brown 1974) that ideology is organized uniformly around a liberal-conservative dimension. And problems with the use of standardized correlation coefficients (see Blalock 1967; Duncan 1975, chap. 4) have been addressed with Barton and Parson's (1977) innovative measure of attitudinal structure and Judd and Milburn's (1980) application of the LISREL model (Jöreskog and Sörbom 1978). The multiple conclusions from all this research are conflicting: support for Converse's notions was claimed by Barton and Parsons (1977), Bishop (1976), Kritzer (1978), Marcus et al. (1974), and Stimson (1975); while results disconfirming Converse were presented by Bennett (1973), Brown (1974), Judd and Milburn (1980), Luttbeg (1968), and, on the stability question at least, by Bishop et al. (1980).

3. Several critics have defined the random component in political attitude scores as something other than the nonopinion which Converse claims it to reveal. This redefinition is explicit in Pierce and Rose's (1974) discussion of the ubiquitous operation of short-term random forces on true attitudes. In fact, Achen's (1975) discussion of instrument vagueness[2] and Judd and Milburn's (1980) reliance on LISREL parameter estimates "corrected" for unreliability also entail such redefinition, as noted in rejoinders to these papers.

Where does the debate stand now? Despite some dissent (e.g., Bishop, Tuchfarber, and Oldendick 1978; Sullivan, Piereson, and Marcus 1978), the historical specificity argument has been accepted, even by Converse (Converse and Marcus 1979), although this is not really a concession because an increase in public attitudinal constraint does not challenge

---

[2] Achen presented his method of estimating reliability as a deviation from the Wiley and Wiley (1970) technique, based on different assumptions and useful for different kinds of data (Achen, p. 1223, n. 16). Readers interested in performing the necessary algebraic manipulations can verify that the Achen technique is an exact algebraic equivalent of the widely used Wiley and Wiley procedure.

the fundamental Converse analysis. Elite-mass comparisons and redefinition of the error component in the data are often intended as challenges to the Converse conclusions, but Converse denies the relevance of either strategy.

One stumbling block to consensus has been the imprecise representation of the Converse analysis. Blurred distinctions between ideological thinking and attitudinal constraint, and between Converse's hypotheses about the respective determinants of these two phenomena, frequently mar the writings of Converse's critics. But the overriding problem with previous examinations of the Converse hypotheses is a mismatch of theory and method. Particularly in his black-and-white model, Converse suggests carving up attitude data in a fashion that diverges radically from conventional measurement theory. Conventional theory acknowledges the existence of true score and error in the data by representing the typical score as uniformly composed of these two elements. Techniques that operate on variance-covariance matrices are bound necessarily to the assumption that observed score composition is uniform across respondents. These techniques can acknowledge heterogeneous contributions of error to observed score variance only when respondents have been divided into subgroups on some a priori basis and subjected to separate subgroup analyses. Converse's analysis is simply incompatible with these constraints. He suggests accounting for true score and error in aggregated data by acknowledging that there are multiple classes of respondents, definable only in terms of response patterns, not on the basis of any a priori criterion. The troublesome but incontrovertible implication is that the techniques typically applied to Converse's analysis are incapable of testing it.

If Converse's analysis is a theory in search of a method, latent class analysis is in some respects a method in search of theories. This currently developing technique has yet to be disseminated widely enough for its potential substantive applications to be explored fully. The task before us is to bring theory and method together.

THE PRESENT STUDY

The Method

The concept of latent class analysis had its genesis in the work of Paul Lazarsfeld (1950). Solutions to computational problems have been provided recently by Leo Goodman (1974a, 1974b), and a Maximum Likelihood Latent Structure Analysis (MLLSA) program implementing these innovations developed by Clifford Clogg (1977). Technical expositions of latent class analysis are available elsewhere (Lazarsfeld and Henry 1968; Goodman 1974a, 1974b, 1975; Clogg 1977, 1979, 1981a, 1981b, 1983)

and will not be summarized here. Rather, this introductory section provides a nontechnical overview of the technique, emphasizing those features that make it uniquely applicable to the Converse propositions.

The basic goal of latent class analysis is to characterize an unobserved categorical variable which "explains" relationships among a set of manifest categorical variables. In terms of its composition and role, the latent variable is closely analogous to the "test factor" in conventional elaboration procedures (Rosenberg 1968), i.e., the ideal latent variable consists of mutually exclusive and exhaustive latent classes within each of which the manifest variables are mutually independent. This requirement is what Lazarsfeld and Henry call the axiom of local independence (1968, p. 22).

The definition of an *unobserved* explanatory variable makes latent class analysis a close analog of factor analytic techniques. One major difference between these two methodologies is that latent class analysis is applicable to qualitative (dichotomous and polytomous) data, whereas factor analysis presumes interval-level observed variables. A related distinction is that factor analysis, in order to explain the observed relationships, identifies one or more continuous underlying dimensions, whereas the unobserved variable identified by the latent class technique consists of classes that are not necessarily even ordinal. This potentially polymorphous character · of the classes constituting a latent variable gives latent class analysis a flexibility several orders greater than that of conventional factor analytic procedures.

In what sense is an unobserved explanatory variable defined by the latent class technique? For the simplest (unrestricted model) case, the Clogg (1977) MLLSA program used here operates from a cross-classification of respondents on the observed categorical variables, together with researcher specification of the number of latent classes, and generates maximum likelihood estimates of two kinds of quantities: the proportion of cases belonging to each latent class, and the proportional distributions of responses on each observed variable within each latent class. Respondents can be assigned to latent classes on the basis of cell membership in the manifest variable cross-classification, and latent class membership can then be used as a variable in subsequent analyses.

Application of the latent structure technique to specific theoretical questions is facilitated greatly by the possibility of imposing restrictions on the latent class model. Central to this project is the restricted "quasi-latent structure" model, which specifies that within one or several latent classes a particular pattern of observed responses is predicted with certainty, whereas observed variables are predicted to be unrelated within the other latent class or classes. Applied to turnover tables representing multiwave response on a single item, the model can specify one class

certain to agree on all waves of the survey, one class certain to disagree, and a third class whose responses are independent across waves—an exact representation of Converse's black-and-white model of attitude stability. (Similarly, applied to interitem correlations, this kind of restricted model could identify classes certain to agree and to disagree on all items, ideologues pro and con, and remaining classes within which observed responses are unrelated. The usefulness of such a model for the investigation of attitudinal constraint is obvious.) Questions about Guttman scalability can be addressed with quasi-latent structure models as well. Here, in the simplest case, the strategy involves setting restrictions so that a class exists to represent each scale type, and a single remaining class contains respondents who are unscalable in the Guttman sense. The estimated size of the residual class provides assessment of response scalability (see Goodman 1975; Clogg and Sawyer 1981; Dayton and Macready 1980).

The MLLSA capacity to assign respondents to latent classes will be particularly useful here. As previously noted, when Converse's propositions are tested using conventional analysis that assumes a single pattern of relationships among observed variables, a problematic a priori designation of subsamples is necessary. By reversing the order of these steps, using response patterns to identify subsamples whose other characteristics can then be ascertained, latent structure analysis provides the solution that has eluded previous research.

The Data

Data for this project are taken from the two three-wave panel studies of voter attitudes conducted by the University of Michigan Institute for Survey Research. From the 1956, 1958, 1960 panel (a national sample of 1,514 respondents), the eight political attitude items discussed in the original Converse paper will be examined here. Each question was introduced with an explicit No Opinion option. Respondents accepting the question were given five alternatives: strongly agree; agree; not sure, it depends; disagree; strongly disagree. The questions read as follows:

> 1. The government should leave things like electric power and housing for private businessmen to handle.
> 2. The government in Washington ought to see to it that everybody who wants to work can find a job.
> 3. If cities and towns around the country need help to build more schools, the government in Washington ought to give them the money they need.
> 4. If Negroes are not getting fair treatment in jobs and housing, the government should see to it that they do.
> 5. The government in Washington should stay out of the question of whether white and colored children should go to the same school.

6. This country would be better off if we just stayed at home and did not concern ourselves with problems in other parts of the world.

7. The United States should keep soldiers overseas where they can help countries that are against communism.

8. The United States should give economic help to the poorer countries of the world even if those countries can't pay for it.

The 1972, 1974, 1976 panel (a national sample of 1,320 respondents) was the basis of the recent Judd and Milburn (1980) article as well as the Converse and Markus (1979) paper; for purposes of comparability the present paper focuses on the five items included in the Judd and Milburn analysis. For the 1972–76 panel, an explicit No Opinion option was not offered, although each question was followed by the phrase "or haven't you thought much about this?" Respondents accepting the question were asked to locate themselves on a seven-point scale. The questions read as follows:

1. Some people are primarily concerned with doing everything possible to protect the legal rights of those accused of committing crimes. Others feel that it is more important to stop criminal activity even at the risk of reducing the rights of the accused. Where would you place yourself on this scale (1 = protect the rights of accused; 7 = stop crime regardless of rights of accused)?

2. Some people feel that the government in Washington should see to it that every person has a job and a good standard of living. Others think that the government should just let each person get ahead on his own. Where would you place yourself on this scale (1 = government see to job and good standard of living; 7 = government let each person get ahead on his own)?

3. Some people think achieving racial integration of schools is so important that it justifies busing children to schools out of their own neighborhood. Others think letting children go to their neighborhood schools is so important that they oppose busing. Where would you place yourself on this scale (1 = bus to achieve integration; 7 = keep children in neighborhood schools)?

4. Some people feel that the government in Washington should make every possible effort to improve the social and economic position of blacks and other minority groups. Others feel that the government should not make any special effort to help minorities because they should help themselves. Where would you place yourself on this scale (1 = government should help minority groups; 7 = minority groups should help themselves)?

5. We hear a lot of talk these days about liberals and conservatives. Here is a seven-point scale on which the political views that people might hold are arranged from extremely liberal to extremely conservative. Where would you place yourself on this scale (1 = extremely liberal; 7 = extremely conservative)?

In the five items from the second panel, the government-interventionist position was always offered first. On the earlier panel, the prointerven-

tionist stance was represented by agreement with five of the statements, but by disagreement with three: utilities and housing, school desegregation, and isolationism. The wording reversals in the first panel offer an opportunity we will exploit to detect any impact of an acquiescence response set. For ease of discussion, government interventionist positions are designated as "liberal" and noninterventionist positions as "conservative." No substantive claims are intended here. We would not deny the multidimensionality of political attitudes or the shifting meanings of the terms liberal and conservative. In particular, the political context of 1981 makes it worrisome to characterize all interventionist foreign policy, including military aid to stop communism, as liberal. The traditional labels were applied, despite these problems, only to simplify an already intricate discussion.

For all but a small set of supplementary analyses, items were dichotomized, dropping the middle category.

The Questions

Analyses reported here address the following questions:

1. How well does Converse's black-and-white model of attitude stability fit dichotomized responses to the eight attitude items from the 1956, 1958, 1960 panel and the five items from the 1972, 1974, 1976 panel? What proportions of respondents are estimated to be in the opinion and nonopinion groups on each issue?

2. Does the introduction of latent classes representing a third force of true score changers improve the fit of the model? What proportion of respondents are estimated to be true score changers? How does allowance for true score change affect the estimated proportions of stable opinion holders and nonopinion respondents?

3. What do the conditional marginal proportions in the nonopinion class imply for the interpretation of nonopinion?

4. Do conclusions change when quasi-latent class models are applied to trichotomized data which include the middle group of "not sure, it depends" responders?

5. How much overlap exists among the various "issue publics"? Are subsets of items Guttman scalable in terms of the proportions of respondents estimated to hold opinions rather than nonopinions?

6. Do opinion holders and nonopinion respondents differ in education or political involvement?

ANALYSES AND FINDINGS

The Black-and-White Model

The black-and-white model, devised by Converse to describe the three-wave responses on the utilities and housing item, is represented by a

quasi-latent structure model with three classes—two classes of perfectly stable opinion holders, those with a 1.0 probability of liberal response at all three waves and those with a 0.0 probability of liberal response at all three waves, and a class of "nonopinion" respondents whose three waves of responses are mutually independent. Estimates of proportional latent class membership take into account the fact that some nonopinion respondents will answer consistently by chance. For each of the eight attitude items from the first panel and the five from the second, the black-and-white model was applied to dichotomized responses of those taking a position at all three waves.

Results are summarized in table 1. Model fit can be assessed using either the likelihood ratio (L^2) or Pearson goodness-of-fit χ^2 statistic, with degrees of freedom equal to the difference between (a) one less than the number of cells in the observed cross-classification and (b) the number of unconstrained parameter estimates. The likelihood ratio χ^2 is used here; for all reported analyses, values of the Pearson χ^2 were negligibly different. The first column in table 1 reports χ^2s for the 4 degrees of freedom (df) "independence" model, which can be viewed as a one-class model with cell proportion predictions based on observed marginal distributions. This serves as a baseline against which to assess improvement in fit when the model acknowledges multiple latent classes of respondents. The second column reports the χ^2 for the black-and-white model itself, with df reduced by 2 because this model estimates two unconstrained latent class membership proportions (the third is constrained so that the proportions sum to 1), along with the conditional probability of liberal response on each of the three manifest variables within the nonopinion latent class (the probability of conservative response is again constrained to bring the sum to 1).

The next three columns present the estimates of proportional membership in the three latent classes. (Estimates of conditional marginal proportions are not presented in table 1 because they are peripheral to the questions at hand.)

The sixth column reports the number of respondents represented in each analysis, and also this N as a percentage of those for whom data were available at all three waves. (The great majority of No Answer respondents excluded from the base for these percentages were apparently unavailable for interviews.) The column 6 percentages are a necessary supplement to the latent class membership estimates: the group excluded from the latent class analysis because of "don't know," "no opinion," or middle category responses on at least one wave of the survey is sizable in all cases and represents the majority on some items. The final column in table 1 synthesizes latent class membership proportions with the in-

TABLE 1

SIGNIFICANCE TESTS AND PARAMETER ESTIMATES FOR THE BLACK-AND-WHITE MODEL OF ATTITUDE STABILITY
APPLIED TO POLITICAL ATTITUDE ITEMS FROM THE 1956, 1958, 1960 AND 1972, 1974, 1976 PANELS

	L^2 (4 df) Independence Model	L^2 (2 df) Black-and-White Model	Liberal Opinion Holders (%)	Conservative Opinion Holders (%)	Nonopinion Respondents (%)	N^a	Opinion Holding in Population[b] (%)
1956, 1958, 1960 items:							
Utilities and housing	161.8	1.0	8.9	35.5	55.5	429 (42)	18.6
Guaranteed jobs	342.1	3.8	41.3	10.5	48.2	697 (64)	33.2
Aid to education	331.8	.0	54.5	8.1	37.4	642 (59)	36.9
Jobs/housing for blacks	294.9	.9	59.7	5.9	34.4	653 (60)	39.4
School desegregation	398.2	10.4**	20.1	26.3	53.7	697 (64)	29.7
Isolationism	262.8	12.2**	45.0	7.1	47.8	704 (65)	33.9
Foreign military aid	88.5	2.7	55.2	2.2	42.6	537 (49)	28.1
Foreign economic aid	159.9	6.9*	36.8	8.7	54.6	446 (41)	18.7
1972, 1974, 1976 items:							
Criminal rights	369.1	3.4	21.1	31.6	47.2	541 (42)	22.1
Guaranteed jobs	350.7	.6	17.4	42.8	39.8	458 (35)	21.1
Busing	289.6	4.6	2.7	83.0	14.4	898 (69)	59.1
Jobs for blacks	418.2	.9	23.3	39.0	37.7	457 (35)	21.8
Liberal/conservative self-identification	490.8	4.8	30.7	46.1	23.1	340 (26)	20.0

[a] Numbers in parentheses represent this N as a percentage of respondents for whom data were available at all three waves.
[b] Percentages were computed by adding the estimated proportions of liberal and conservative opinion holders (from cols. 3 and 4), multiplying this sum by the proportion of the responding sample represented in the latent class analysis (from col. 6), and multiplying that product by 100.
* Significant at the .05 level.
** Significant at the .01 level.

formation presented in column 6 and represents opinion holders as a proportion of all survey respondents from whom data was available, thus estimating the rate of opinion holding in the population.

Table 1 indicates that the black-and-white model of attitude stability fits the three waves of responses to most of these items remarkably well. The χ^2 assessing divergence of observed cell frequencies from model predictions reaches the .05 level of significance only for the school desegregation, isolationism, and foreign economic aid items, and even for these the black-and-white model χ^2 is dramatically smaller than the baseline model χ^2.[3] The utilities and housing item is particularly well fitted by the black-and-white model, as Converse had suggested, but on four other issues the model provides a similarly good fit to the data. In terms of model fit, there seems to be no systematic difference between the two panels.

Turning to the estimated proportions in the nonopinion latent class, we see that the utilities and housing item indeed exceeds all others, but the nonopinion estimates on school desegregation and foreign economic aid are only negligibly smaller. Foreign economic aid is the sort of remote issue which Converse would predict to elicit a substantial number of nonopinion responses, but school desegregation is not. The column 6 percentages of respondents taking a position at all three waves suggest that school desegregation may be a different kind of issue than the other two. On the utilities and housing and foreign economic aid issues, the substantial nonopinion response is accompanied by low response rates, plausibly signaling the lack of interest and/or information that Converse discussed. But school desegregation, which shows both substantial nonopinion response and a high response rate, may be an item which elicits nonopinion for other reasons, either because respondents are ambivalent or because they feel compelled to take a position even when they do not have one.[4]

In general, proportional membership in the nonopinion latent class is

[3] The difference between likelihood ratio χ^2 statistics for two nested latent class models is not necessarily distributed as χ^2 when the restrictions to be tested amount to putting parameters at "boundaries" like 0 or 1 (see Haberman 1981). Thus the difference between baseline and black-and-white model χ^2s cannot be used to test the significance of parameters differentiating the two models, since the baseline model is obtained from the black-and-white model by setting certain latent class proportions at 0. However, it is only the χ^2 *difference* that presents problems for *conditional* comparisons. The χ^2 statistics for the black-and-white model can be used to assess this model in unconditional terms.

[4] A complementary point may be made about the two issues with the *lowest* proportions of nonopinion respondents, busing and liberal/conservative self-identification. Busing, which shows a high response rate along with a small nonopinion class, is apparently an issue on which most Americans have stable opinions. In contrast, most respondents decline to commit themselves on the liberal/conservative self-identification question, and the latent class analysis showing predominantly stable opinion holding on this issue represented a very select group of respondents in the first place.

smaller for the second panel, although the usefulness of this comparison is limited: not only did the no opinion filter differ for the two panels, but there was no exact replication of questions across the two panels and only two issues were represented on both—questions of guaranteed jobs and jobs for blacks. Converse and Markus's (1980) conclusion that attitudinal stability on these two issues was similar across the panels is only partially supported by the present results: nonopinion response on the guaranteed jobs item did decline from the 1956–60 panel to the 1972–76 panel.

Finally, the estimates of opinion holding in the population indicate that Converse was right on the mark in asserting that less than 20% of the population had an opinion on the utilities and housing issue—18.6% is the estimate here. Opinion holding on the foreign economic aid question is equally low. Other percentages for the 1956–60 panel range between 28% and 40%, with the jobs/housing for blacks question revealing the highest proportion of opinion holding. Percentage estimates for the 1972–76 panel are generally lower, in the 20%–22% range for four of the items. Again a race item reveals the greatest proportion of opinion holding— the busing question with a striking estimate of 59%. As Converse had noted, the generally low level of these estimates should give pause to students of political affairs. The relatively high proportion of opinion holding on race issues is noteworthy as well, a testimony to the salience of race in American political thought.[5]

The Third Force

Although the fit of the black-and-white model to many of the items considered here could hardly be improved, we also operationalized Converse's contention that on many issues a third force exists, holding actual but changing opinions. The latent structure model to represent this idea contains the three classes comprising the black-and-white model and two others: a becoming-liberal class, defined by conditional marginal probabilities of a liberal response constrained to be 1.0 for the first-wave data

[5] MLLSA is not the only program with the capacity to perform latent class analysis: Haberman's (1979) LAT program is also applicable to the quasi-latent structure models examined here. A major difference between the two programs is that the MLLSA parameter estimates are proportions (of latent class membership and conditional marginal distributions), whereas LAT uses a regression-type model and estimates main effects and interaction effects for the cross-classification of manifest and latent variables. MLLSA was chosen for use here because its output translates so directly into the terms of Converse's discussion. But the central black-and-white model analyses were replicated using the LAT program. Significance statistics provided by the two programs were identical, and calculations using LAT effect estimates yielded the same latent class membership proportions as those estimated by MLLSA. Unfortunately, standard error estimates provided for the LAT effects cannot be used readily to derive standard errors for the proportions presented in table 1. However, LAT standard error and effect estimates for these analyses may be obtained from me.

and 0.0 for the third-wave data; and a becoming-conservative class, defined by first- and third-wave conditional marginal probabilities constrained in the symmetrically opposite direction (0.0 and 1.0, respectively). A further divergence from the black-and-white model was necessitated by the requirement that an identifiable latent class model, that is, a model that produces unique parameter estimates, can have no more nonredundant parameters than the number of cells in the manifest variable cross-classification minus 1.[6] In order to meet this requirement when the five-class model was applied to the $2 \times 2 \times 2$ turnover tables, an invariance constraint was imposed on the conditional marginal proportions within the nonopinion class, implying identical proportions of liberal response at all three waves of the survey. This five-class model was applied to attitude items from the first and second panels, and parameter estimates were examined.

Although the five-class model described above is identifiable when applied to an eight-cell manifest variable cross-classification, there must be one fewer nonredundant parameter, that is, at least 1 df, for the model to be testable. Fortuitously, in each of the 13 five-class analyses, the data suggested model simplification that achieved the necessary degree of freedom. In some cases, the estimated membership in one latent class approached 0. In other cases, the conditional marginal proportions for the second-wave responses among one of the third-force classes approached or reached the extreme values of 1.0 or 0.0 (indicating, e.g., that all change took place between the first and second waves of the survey). In no case did the reduced model produce parameter estimates substantially different from those of the full five-class model. Table 2 presents the results of analyses using the reduced models suggested by the original five-class analysis for each particular item and yielding results negligibly different from it.

The first column in table 2 is again a baseline likelihood ratio χ^2, this time for a 4 df black-and-white model with an invariance restriction imposed on conditional marginal proportions within the nonopinion class. Although we will ultimately compare the five-class model with the black-and-white model represented in table 1, the baseline χ^2 for the more heavily restricted three-class model allows us to see the effects of introducing the two third-force classes per se, unconfounded by effects of imposing the invariance restriction on conditional marginal proportions. (The baseline χ^2s in table 2 indicate that the invariance restriction does

---

[6] This condition is necessary but not sufficient for identification. Model identifiability also requires that the matrix of derivatives of nonredundant parameters with respect to nonredundant manifest variable cross-classification cell proportions be of full column rank. Some models have few enough nonredundant parameters to be testable but are, nonetheless, not fully identified. For details, see Goodman (1974b).

TABLE 2

SIGNIFICANCE TESTS AND PARAMETER ESTIMATES FOR THE BLACK-AND-WHITE AND THIRD-FORCE MODELS APPLIED TO POLITICAL ATTITUDE ITEMS FROM THE 1956, 1958, 1960 AND 1972, 1974, 1976 PANELS

	L^2 (4 df) Baseline Model	L^2 (df in parentheses) Third-Force Model	Liberal Opinion Holders	Conservative Opinion Holders	Nonopinion Respondents[a]	Becoming Liberal[b]	Becoming Conservative[c]
1956, 1958, 1960 panel:							
Utilities and housing	4.1	2.9 (2)	8.6	37.0	52.4 (.43)		2.0 (.95)
Guaranteed jobs	16.7**	1.6 (1)	37.0	12.9	42.6 (.67)	5.7 (.20)	1.7 (1.00)
Aid to education	33.3**	6.0* (2)	57.1	8.2	26.9 (.58)		7.8 (.31)
Jobs/housing for blacks	7.7	.9 (1)	59.9	6.4	29.6 (.64)	3.8 (.55)	.3 (1.00)
School desegregation	11.3*	.1 (1)	22.5	26.3	44.9 (.53)	3.9 (1.00)	2.4 (.82)
Isolationism	16.8**	6.7**(1)	44.4	7.8	42.7 (.71)	2.8 (.90)	2.3 (1.00)
Foreign military aid	15.0**	.1 (2)	61.4	2.2	30.6 (.73)	5.8 (.85)	...
Foreign economic aid	20.1**	.1 (2)	43.4	7.6	41.9 (.56)	8.1 (.94)	...
1972, 1974, 1976 panel:							
Criminal rights	15.1**	2.2 (2)	23.3	29.9	41.3 (.42)		5.5 (.15)
Guaranteed jobs	3.5	.2 (2)	17.5	43.9	35.3 (.45)		3.3 (.67)
Busing	4.8	2.4 (1)	2.7	84.9	10.7 (.28)	.9 (.0)	.8 (.67)
Jobs for blacks	11.2**	4.6* (1)	23.1	41.3	29.8 (.47)	.9 (.0)	4.8 (.68)
Liberal/conservative self-identification	15.6**	4.4** (1)	32.2	37.3	23.3 (.18)	2.3 (1.0)	4.9 (.33)

a Numbers in parentheses are conditional probabilities of liberal response.
b Numbers in parentheses are the probabilities that the shift to liberal opinion occurred between the first and second waves of the survey.
c Numbers in parentheses are the probabilities that the shift to conservative opinion occurred between the first and second waves of the survey.
* Significant at the .05 level.
** Significant at the .01 level.

some violence to the data. Most are significant beyond the .01 level, whereas table 1 had shown a comparable divergence from the simple black-and-white model for only two of the 13 items. Conditional marginal distributions for the nonopinion class did, in fact, shift over time when left unconstrained as in the simple black-and-white model.)

Comparing the χ^2 values for the reduced five-class models with the baseline χ^2s, we see that introducing third-force classes of opinion changers always improved on the fit of the baseline model. However, our real interest is in the comparison between the simple black-and-white model of table 1 and the reduced five-class models. Of the three items that diverged significantly from table 1 black-and-white models predictions, two (school desegregation and foreign economic aid) are better fitted by the reduced five-class model: χ^2s are reduced to a trivial value. Isolationism, the third item which diverged significantly from the simple black-and-white model, shows a significant χ^2 for the reduced five-class model as well. (As suggested in the earlier discussion, where the data seem to be *less* adequately fitted by the reduced five-class model than by the simple black-and-white model, table 2 baseline χ^2s indicate that the explanation lies in the invariance restriction on conditional marginal proportions.) In short, the model acknowledging true score change did represent the data better for two of the survey questions, but, on balance, both the simple black-and-white model and the five-class third-force model fitted these opinion data exceptionally well.

Comparing table 1 and table 2 estimates of proportional membership in the nonopinion class, we see that introduction of the third-force classes does bring a reduction for most of the items—for some, a drop of more than 10 percentage points. Estimates of nonopinion change least from table 1 to table 2 for the utilities and housing and the liberal/conservative self-identification questions, and in table 2 the utilities and housing item stands out more clearly as the foremost eliciter of nonopinion response. But estimated proportions of nonopinion response remain substantial for all of these issues.

Estimates of proportional membership in the third-force classes, though small, are substantively interesting. For example, it appears that 1956–60 attitude change on foreign military and economic aid was uniformly toward the liberal (interventionist) position. The preponderant shift in the 1956–60 panel was in the liberal direction on the race and guaranteed jobs items, while a conservative shift was dominant on the question of aid to education. Results for the 1972–76 panel show uniform change toward the conservative position, except for the busing question on which the great majority of respondents are estimated to be stable conservatives and on which virtually no attitude change was observed. The proportions

in parentheses in the last two columns of table 2 indicate the typical timing of attitude change. For example, although few respondents are estimated to have changed their attitudes on the 1972, 1974, 1976 race items, the conditional marginal proportions for second-wave responses suggest that the rare conservative shifts were most likely to have occurred between 1974 and 1976.

A final point of interest in table 2 is the conditional probabilities of liberal response among the nonopinion respondents, presented in parentheses in the third column from the right. The pattern of these proportions has important implications for the interpretation of nonopinion. The strongest definition of nonopinion, that of literally random response, would be reflected in proportions clustered tightly around the .5 level. Converse (1974) has denied explicitly any intention to suggest equiprobable responses among nonopinion people, but others have assumed this definition, and it is noteworthy that our data do not support the equiprobability idea.

A plausible hypothesis is that the divergence from equiprobable responses among the nonopinion class reveals some influence of item form rather than content, for example, an acquiescence response set. If this were the case, the 1956–60 panel nonopinion groups would show greater liberal response on the five items where agreement indicated liberalism than on the three items where agreement meant taking a conservative position. The proportions presented in table 2 do not reveal this pattern either.

The pattern which table 2 does reveal is that conditional marginal proportions of the nonopinion class mirror the proportional distributions of opinion holders on that particular issue. Where there is a great preponderance of liberal respondents among the stable opinion holders, there is a similar preponderance of liberal response in the nonopinion class. Table 2 estimates were used to rank the 13 items in terms of the ratio of liberal to conservative opinion holders and also in terms of the proportion of liberal responses within the nonopinion class; the Spearman ρ is .79. A similar point can be made about patterns of change over time. For items on which table 2 shows a preponderance of true score change in one direction, unconstrained conditional marginal proportions of the simple black-and-white model reveal that nonopinion class response distributions changed over time in a parallel direction. (In fact, it is these items that have a particularly large table 2 baseline χ^2, precisely because a shift over time in the response distributions of nonopinion respondents was denied in the invariant distribution constraint.)

These observations suggest that nonopinion must be described carefully. Within the sizable nonopinion class on each of these issues, responses

at one wave of the survey are indeed independent of responses at the other two waves. However, this nonopinion group is *not* producing answers which are independent of item content. They tend to be liberal on those issues on which opinion holders are liberal, conservative on issues where conservative opinion holders are in the majority. Although their own attitudes are not crystallized well enough to produce consistent responses, the nonopinion class is apparently aware enough of the considerations that lead opinion holders to their positions (or perhaps aware enough of the dominant sentiment among the opinionated) to mirror these in their own collective response.

Exploratory Analysis of Trichotomized Responses

To supplement the analyses described earlier, latent class models were applied to trichotomized data on the utilities and housing item, including equivocal, middle-category responses as the additional manifest variable level.

The data were well fitted by a four-class analog to the simple black-and-white model (the additional class providing for consistent "not sure, it depends" respondents), but there were estimated to be virtually no stable "not sure, it depends" respondents (.2% of the total). Proportional membership estimates for the other three latent classes differ from the estimates of the simple black-and-white model by almost the exact amount that would be predicted if the "not sure, it depends" respondents had simply been added to the nonopinion class.

A three-class model was also applied, imposing the restrictions Clogg (1979) has suggested for Likert-type data: conservative responses were prohibited for one class, liberal responses were prohibited for a second class, and the middle class had unrestricted conditional marginal proportions. The data were well fitted by this model, and conditional marginal proportions were congruent with the interpretation of three ordered classes. However, the estimates of latent class membership were very close to those produced by the trichotomous data version of the black-and-white model: the 66% nonopinion figure identified by the black-and-white model compares closely to the 61% estimated to fall in the middle, fence-sitting category of the "three-ordered-classes" model.

In sum, an elaborated version of the black-and-white model fitted these data well. A three-ordered-classes model fitted the data equally well but suggested the same substantive conclusion as the black-and-white model—that most respondents have neither a liberal nor a conservative opinion on this issue. The Converse conclusion holds up well under this examination. (Additional data analysis, using latent class techniques that rep-

resent more conventional measurement models, is described in the Appendix.)

The Overlap of Issue Publics

As noted earlier, Converse concluded that there is only modest overlap of opinion holding across issues, and that issues cannot be ordered so as to reveal an expanding core of stalwart opinion holders. The quasi-latent structure model capable of testing this conclusion specifies the existence of one latent class representing each Guttman scale type (within each of these, cases have certain probability of representing the prescribed pattern of values), and a final class of "unscalable" respondents for whom status on the several manifest variables is independent.

Three analyses were used to assess the overlap of issue publics: one included the three foreign policy issues from the 1956–60 panel ($N = 243$); a second ($N = 174$) included four of the five domestic policy issues from the 1956–60 panel (small cell frequencies precluded analysis on the jobs/housing for blacks issue); and a third ($N = 153$) included three of the five 1972–76 items (small cell frequencies prohibited analysis on the busing and liberal/conservative self-identification questions). For each included item, the latent structure assignment of respondents to latent classes was used to create a dichotomous variable, representing opinion holding versus nonopinion response. The three analyses were then based on cross-classifications of respondents by opinion-holding status on the issues represented. Marginal distributions were used to order the items and determine the Guttman scale types.

For the 1956–60 foreign policy items, L^2 for the 4 df independence model was 20.9, suggesting that opinion-holding status on the three survey questions is not entirely independent. The 0 df model which included all four scale types and an unscalable class yielded nontrivial latent class membership proportions only for the class representing opinion holders on all three issues and the class of unscalable respondents. A reduced, 3 df model representing these two classes fitted the data extremely well ($L^2 = 1.3$). The analysis estimated 25.4% of the respondents to be stable opinion holders on all three issues; opinion-holding status on the remaining 74.6% was unpredictable from one issue to another.

Examination of the four 1956–60 domestic policy items yielded a non-significant χ^2 value of 7.5 for the 11 df independence model; with respect to guaranteed jobs, utilities and housing, school desegregation, and aid to education, response stability is independent from one issue to the next.

Turning to the guaranteed jobs, jobs for blacks, and criminal rights questions on the second panel, χ^2 for the 4 df independence model attained the barely significant value of 11.7. Analysis using the full model revealed

near-zero membership proportions for all but the class of stable opinion holders on all three issues and the latent unscalable class. A reduced, two-class, 3 *df* model representing these two groups provided a good fit ($L^2 = 2.8$) and produced estimates that 25.9% of the respondents are opinion holders on all three issues; for the other 74.1%, stable opinion holding is unpredictable across issues.

To summarize, Converse was apparently right in concluding that there is relatively little overlap among issue publics. For two of the three item subsets examined here, a respectable minority does hold stable opinions on all issues. But for most respondents on these items, and for all respondents on the 1956–60 domestic policy items, the holding of stable opinion on one issue implies nothing about opinion-holding status on other questions.

The Educational Level and Political Involvement of Opinion Holders and Nonopinion Respondents

As mentioned earlier, Converse explicitly disavowed education as a predictor of opinion holding and noted that partisan political activism, though the best single predictor, is itself inadequate. However, analysts responding to Converse have often assumed that opinion holders and nonopinion respondents can be differentiated on the basis of education or political involvement, and many attempts to test Converse have entailed comparison of subsamples that differ on these variables. Our earlier finding that opinion holding is largely issue specific demonstrates that such comparisons are misguided. The analyses reported below reinforce this point.

The black-and-white model assignments of respondents to latent classes were used to compare opinion holders and nonopinion respondents on each issue in terms of education and three measures of political involvement—self-reported tendency to follow political affairs on a daily basis, self-reported history of attempting to influence the voting behavior of others, and (for the 1972–76 panel only) interviewer ratings of respondents' level of political information. Table 3 summarizes the results.

Looking first at data from the 1956–60 panel, we see that only some issues show education differences between opinion holders and nonopinion respondents, and the differences that do exist are inconsistent in direction. Opinion-holding status shows even weaker relationships to the political involvement variables, especially the "follow politics closely" measure. The direction of the differences in political involvement tends to parallel the direction of the education effects, suggesting that all these variables may serve to index social status.

The pattern of differences between opinion holders and nonopinion respondents across the various issues supports the plausible notion that

TABLE 3

EDUCATION AND POLITICAL INVOLVEMENT AMONG OPINION HOLDERS AND NONOPINION RESPONDENTS

	SOME COLLEGE (%)		FOLLOW POLITICS (%)		TRY TO INFLUENCE VOTE (%)		RATED HIGH ON INFORMATION (%)	
	Opinion Holders	Nonopinion Respondents	Opinion Holders	Nonopinion Respondents	Opinion Holders	Nonopinion Respondents	Opinion Holders	Nonopinion Respondents
1956, 1958, 1960 panel:								
Utilities and housing	29.3	25.9	23.2	21.1	38.6	39.4	...	...
Guaranteed jobs	16.9	23.7	18.9	18.1	29.4	33.8	...	...
Aid to education	15.9	26.9	19.5	24.0	31.6	36.8	...	...
Jobs/housing for blacks	18.4	21.9	20.5	20.5	32.1	34.2	...	...
School desegregation	22.4	22.7	22.3	21.0	35.9	31.4	...	...
Isolationism	29.0	14.5	22.4	20.6	37.2	34.6	...	...
Foreign military aid	21.6	23.8	21.4	21.6	36.6	32.0	...	...
Foreign economic aid	26.4	17.6	24.2	17.7	35.4	33.1	...	...
1972, 1974, 1976 panel:								
Criminal rights	42.5	40.1	44.5	42.5	43.7	38.9	48.8	39.1
Guaranteed jobs	46.7	41.4	49.5	38.5	45.2	45.2	49.7	50.4
Busing	37.5	32.9	43.8	31.1	37.5	36.5	43.7	32.9
Jobs for blacks	44.8	39.4	40.1	45.3	43.8	41.4	46.9	45.3
Liberal/conservative self-identification	63.6	59.3	56.9	49.2	55.5	57.6	66.1	61.0

people develop crystallized attitudes on issues that have personal relevance. On the questions of guaranteed employment, federal aid to schools in poorer communities, jobs and housing for blacks, and sending troops overseas, stable opinion holders tend to be the less educated respondents. And indeed, lower SES groups do feel the impact of such decisions most immediately. Opinion holders on the utilities and housing question and the other two foreign policy issues tend to be the more highly educated respondents. These questions seem to have particular relevance for the high SES groups who have the greater stake in existing economic relations.

Results for the second panel are somewhat different. Education differences are small but consistently favor the opinion holders. Again, the political involvement variables generally parallel the education measure. (The "follow politics closely" item differentiates opinion holders from nonopinion respondents more sharply in this panel, but a change in item wording may account for this shift.) The change in political and economic climate between the 1950s and the 1970s may be an explanation for the differing pattern of results in the two panels. During recent years, public debate about federal intervention to aid less privileged groups (represented in the guaranteed jobs and jobs for blacks items) has focused less on the need of the beneficiaries than on the sacrifice which such efforts impose on the relatively privileged. In the 1970s, higher SES groups may have come to see these issues as personally relevant in a way that the high SES respondents of the 1950s did not.

Whatever the explanation of these particular patterns, it is clear that opinion holders and nonopinion respondents are not sharply or consistently distinguished in terms of either education or political involvement.[7]

SUMMARY AND CONCLUSIONS

Converse's black-and-white model of attitude stability—positing the existence of stable opinion holders and a group of "nonopinion" respondents whose positions vary unpredictably over time—is an apt description of responses to political attitude questions in both the 1956, 1958, 1960 and

[7] Two additional findings might be noted. First, on many issues both the opinion holders and the nonopinion respondents show higher levels of education and political involvement than the average for the total sample of respondents. The less educated and politically involved did often decline to take a position and thus are represented neither as opinion holders nor as nonopinion respondents. Second, on many issues liberal and conservative opinion holders differed dramatically in terms of education and political involvement. More often than not, the nonopinion respondents showed educational and political involvement levels that were intermediate between those of the consistent liberals and those of the consistent conservatives. It is indeed much too simple to characterize the nonopinion groups as relatively uneducated or uninformed.

the 1972, 1974, 1976 American voter panel studies. Estimated proportions of nonopinion respondents differ from item to item, but almost all are substantial. There was no indication of a general increase in stability from the 1950s to the 1970s. Where nonopinion response was combined with equivocal answers and nonresponse, rates of opinion holding in the population were estimated. These ranged from a low of 19% for the utilities and housing item which had been the focus of Converse's writing to an atypical high of 59% for the busing item in the second panel.

A latent structure model acknowledging true score change also fitted the data quite well and typically estimated a somewhat smaller proportion of nonopinion respondents than that of the simple black-and-white model. The third force of true score changers was always estimated to be small, however, and the model allowing for true score change still identified sizable nonopinion classes.

Inclusion of equivocal or middle-category responses did not change these conclusions—respondents who gave at least one equivocal response simply swelled the ranks of the nonopinion class.

Estimates of manifest variable response distributions within the nonopinion class have important implications for the interpretation of nonopinion. Though the responses of this group were independent from one wave to the next, the marginal distributions of their responses paralleled the responses of opinion holders. Apparently the nonopinion group is responsive enough to item content that their aggregated responses reflect the climate of opinion among those who whose attitudes are better crystallized.

As Converse had concluded, there is relatively little overlap of issue publics. Those who hold opinions on one issue will not necessarily be opinion holders on other questions. Opinion holders and nonopinion respondents do not differ consistently in education or measured political involvement. Relationships of education and political involvement to opinion holding suggest that respondents hold stable opinions on issues that have greatest personal relevance. These conclusions are congruent with the Converse analysis. And they underline the futility of testing Converse's ideas by comparing attitudinal stability in groups high and low on education or political involvement.[8]

Conclusions drawn in this study may serve to dispel the normative overtones that have clouded debate over the Converse analysis. If it is

[8] Schuman and Presser's recent (1981) book considers many topics touched on in this paper (e.g., characteristics of those who decline to offer responses to survey questions, the meanings of middle-category responses, and definitions of opinion and nonopinion). Because the approach of these analysts is quite different from that of Converse—they focus on effects of item form rather than on patterns of response stability—their conclusions have not been summarized here. Readers seeking a comprehensive understanding of influences on survey responses will find the intricate Schuman and Presser discussion a valuable resource.

acknowledged that nonopinion response is not correlated consistently with social status, that it is issue specific rather than a characteristic of certain respondents, and that it does not represent complete unresponsiveness to item content, nonopinion can be relieved of its stigma as an elitist notion and treated as a challenging problem for political attitude research.

APPENDIX

Although this paper has focused on the quasi-latent structure models that represent Converse's opinion and nonopinion hypothesis, the following paragraphs illustrate that latent class techniques can also be used to represent more conventional measurement models.

The model conventionally applied to multiwave data depicts true scores which change over time and are imperfectly reflected in responses at each wave. This model is too elaborate for a $2 \times 2 \times 2$ cross-classification of dichotomized manifest variable responses, but a simplified version of this model can be applied—a two-class model which represents a single underlying variable (a stable true score), imperfectly represented in the three waves of measured responses. An unrestricted two-class model of this form is identified but not testable for three-wave dichotomized data; however, introducing a reasonable invariance constraint on conditional marginal proportions, requiring that within each latent class the probability of a given response be the same at all three waves of the survey, yields a 4 *df* model which is testable as well as identified.

For exploratory purposes, this restricted two-class model was fitted to the three-wave data on five of the 13 items examined in this paper: utilities and housing, aid to education, foreign economic aid, jobs for blacks (from the 1972–76 panel), and liberal/conservative self-identification. In each case, the analysis yielded interpretable results, suggesting the existence of one class likely but not certain to give liberal responses and a second class likely but not certain to give conservative responses. However, the data diverged significantly from model predictions for all but the housing and utilities item. On the whole, the data were fitted less well by this version of the conventional measurement model than by the black-and-white model or the third-force model which acknowledged true score change.

A hybrid model, incorporating elements of the black-and-white model and the conventional measurement model, was also applied to three-wave dichotomized responses on these five survey items. The hybrid specified four latent classes—consistent liberal and consistent conservative classes, together with two less heavily restricted classes potentially interpretable as latent liberal and latent conservative classes. Some variants of this model proved to be underidentified, despite adequate degrees of freedom.

A variant which is identified imposes an "equal error rates" restriction, implying that at each wave the proportion of liberal responses within the latent liberal class must equal the proportion of conservative responses within the latent conservative class. This model fitted the data well for all but the liberal/conservative self-identification item. However, the more parsimonious black-and-white model provided an equally good fit for one of these items and a better fit for the other four.

The $3 \times 3 \times 3$ cross-classification of *trichotomized* responses to a survey question does allow application of a latent class model which represents the more elaborate version of conventional measurement theory, allowing for true score change over time. The latent structure representation of this idea is an eight-class model, where the eight classes are envisioned as cells of a $2 \times 2 \times 2$ cross-classification of three interrelated, dichotomous latent variables, one latent variable representing true scores at each wave of the survey. Conditional probabilities of observed variable responses can be restricted so that they differ among some subsets of latent classes but not among others, implying that observed responses at a given wave are directly affected only by true scores at that wave. Estimates of conditional probabilities can be used to calculate the equivalent of a path coefficient relating each wave of manifest responses to the appropriate true score dimension, and the latent class proportions can be used to estimate relationships among the true score dimensions.

Eight-class models of this form were applied to the trichotomized responses on the utilities and housing item. The 7 *df* model described above proved to be underidentified. A set of theoretically reasonable additional restrictions, implying that observed responses at the three waves of the survey were equally influenced by the respective true scores (i.e., that the reliability of the measured attitudes was invariant over time), was also underidentified.

Two other kinds of restrictions, when added to the invariant reliability restriction, successfully produced identified models. The first was a kind of equal error rate restriction, implying equal probability of liberal response within all true liberal classes and of conservative response within all true conservative classes. This 16 *df* model fitted the data well, and estimates of conditional marginal proportions were interpretable; 83% of those in the true liberal class for a particular wave gave liberal responses at that wave (and by necessity the same percentage of true conservatives gave conservative responses). The second kind of additional restriction producing an identified model implied a constant rate of true score change: among those whose true attitudes changed in either direction from wave 1 to wave 3, the change was as likely to have occurred between waves 1 and 2 as between waves 2 and 3. This 17 *df* model also fitted the data well and again yielded interpretable estimates.

These exploratory analyses were described primarily for their methodological interest. Envisioning the eight-class latent structure model as a representative of conventional measurement theory seems to be a particularly useful innovation. It may prove feasible to adjust latent structure analysis programs so that the researcher can specify a lag-1 model, that is, that true scores at a given wave are dependent solely on true scores at the immediately preceding wave. (This would entail constraining latent class membership estimates to imply that wave 1 and wave 3 true score dimensions are independent within levels of wave 2 true scores.)

These analyses also have substantive implications. When applied to the dichotomized data which had been Converse's primary focus, the simplified model of conventional measurement theory did not fit the data as well as the black-and-white model, strengthening support for our earlier conclusions. However, trichotomized responses to the utilities and housing item were well represented both by the trichotomous version of the black-and-white model and by the eight-class representative of the conventional measurement model. Only within limits does empirical evidence provide critical tests of rival theoretical frameworks. Where the data are well fitted by two quite divergent models, the choice between perspectives must be based on other considerations.

REFERENCES

Abelson, R. P., E. Aronson, W. J. McGuire, T. M. Newcomb, and P. H. Tannenbaum, eds. 1968. *Theories of Cognitive Consistency: A Sourcebook.* Chicago: Rand-McNally.

Achen, C. H. 1975. "Mass Political Attitudes and the Survey Response." *American Political Science Review* 69:1218–31.

Barton, A. H., and R. W. Parsons. 1977. "Measuring Belief System Structure." *Public Opinion Quarterly* 41:159–80.

Bennett, S. 1973. "Consistency among the Public's Social Welfare Attitudes in the 1960s." *American Journal of Political Science* 17:544–70.

Bishop, G. D., D. L. Hamilton, and J. B. McConahay. 1980. "Attitudes and Non-attitudes in the Belief System of Mass Publics." *Journal of Social Psychology* 110:53–64.

Bishop, G. F. 1976. "The Effect of Education on Ideological Consistency." *Public Opinion Quarterly* 40:337–48.

Bishop, G. F., A. J. Tuchfarber, and R. W. Oldendick. 1978. "Change in the Structure of American Political Attitudes: The Nagging Question of Question Wording." *American Journal of Political Science* 22:250–69.

Blalock, H. M. 1967. "Path Coefficients versus Regression Coefficients." *American Journal of Sociology* 72:675–76.

Brown, S. R. 1974. "Consistency and the Persistence of Ideology." *Public Opinion Quarterly* 34:60–68.

Campbell, A., P. E. Converse, W. E. Miller, and D. E. Stokes. 1960. *The American Voter.* New York: Wiley.

Clogg, C. C. 1977. "Unrestricted and Restricted Maximum Likelihood Latent Structure Analysis: A Manual for Users." Working Paper 1977–09. University Park: Pennsylvania State University, Population Issues Research Center.

———. 1979. "Some Latent Structure Models for the Analysis of Likert-type Data." *Social Science Research* 8:287–301.

———1981a. "Latent Structure Models of Mobility." *American Journal of Sociology* 86:836–68.

———. 1981b. "New Developments in Latent Structure Analysis." Pp. 214–46 in *Factor Analysis and Measurement in Sociological Research*, edited by D. M. Jackson and E. F. Borgatta. Beverly Hills, Calif.; Sage.

———. 1983. "Some Statistical Models for Analyzing Why Surveys Disagree." In *Surveying Subjective Phenomena*. Vol. 2. Edited by C. F. Turner and E. Martin. New York: Basic, in press.

Clogg, C. C., and D. O. Sawyer. 1981. "A Comparison of Alternative Models for Analyzing the Scalability of Response Patterns." Pp. 240–80 in *Sociological Methodology 1981*, edited by S. Leinhardt. San Francisco: Jossey-Bass.

Converse, P. E. 1964. "The Nature of Belief Systems in Mass Publics." Pp. 206–61 in *Ideology and Discontent*, edited by D. E. Apter. London: Free Press of Glencoe.

———. 1970. "Attitudes and Non-attitudes: Continuation of a Dialogue." Pp. 168–89 in *The Quantitative Analysis of Social Problems*, edited by E. R. Tufte. Reading Mass.: Addison-Wesley.

———. 1974. "Comment: The Status of Nonattitudes." *American Political Science Review* 68:650–66.

———. 1975. "Public Opinion and Voting Behavior." Pp. 75–170 in *Nongovernmental Politics*, edited by F. I. Greenstein and N. W. Bolsby. Reading, Mass.: Addison-Wesley.

———. 1980. "Comment: Rejoinder to Judd and Milburn." *American Sociological Review* 45:644–46.

Converse, P. E., and G. B. Markus. 1979. "Plus ça change . . . the New CPS Election Study Panel." *American Political Science Review* 73:32–49.

Dayton, C. M., and G. B. Macready. 1980. "A Scaling Model with Response Errors and Intrinsically Unscalable Respondents." *Psychometrika* 45:343–58.

Duncan, O. D. 1975. *Introduction to Structural Equation Models*. New York: Academic Press.

Field, J. O., and R. E. Anderson. 1969. "Ideology in the Public's Conceptualization of the 1964 Election." *Public Opinion Quarterly* 33:380–98.

Goodman, L. A. 1974a. "The Analysis of Systems of Qualitative Variables When Some of the Variables Are Unobservable." *American Journal of Sociology* 79:1179–1259.

———. 1974b. "Exploratory Latent-structure Analysis Using Both Identifiable and Un-identifiable Models." *Biometrika* 61:215–31.

———. 1975. "A New Model for Scaling Response Patterns: An Application of the Quasi-independence Concept." *Journal of the American Statistical Association*. 70:755–68.

Haberman, S. J. 1979. *Analysis of Qualitative Data*. Vol. 2. New York: Academic Press.

———. 1981. "Tests for Independence in Two-way Contingency Tables Based on Canonical Correlation and on Linear-by-Linear Interaction." *Annals of Statistics* 9:1178–86.

Jöreskog, K. G., and D. Sörbom. 1978. LISREL: *Analysis of Linear Structural Relationships by the Method of Maximum Likelihood*. Chicago: National Educational Resources.

Judd, C. M., and M. A. Milburn. 1980. "The Structure of Attitude Systems in the General Public: Comparisons of a Structural Equation Model." *American Sociological Review* 45:627–43.

Kritzer, H. M. 1978. "Ideology and American Political Elites." *Public Opinion Quarterly* 42:484–502.

Lazarsfeld, P. F. 1950. "The Logical and Mathematical Foundation of Latent Structure Analysis." Pp. 362–412 in *Measurement and Prediction*, edited by S. A. Stouffer et al. Princeton, N.J.: Princeton University Press.

Lazarsfeld, P. F., and N. W. Henry. 1968. *Latent Structure Analysis*. Boston: Houghton Mifflin.

Luttbeg, N. R. 1968. "The Structure of Beliefs among Leaders and the Public." *Public Opinion Quarterly* 32:398–409.

Marcus, G., D. Tabb, and J. L. Sullivan. 1974. "The Application of Individual Differences Scaling to Measures of Political Ideologies." *American Journal of Political Science* 18:405–20.

Margolis, M. 1977. "From Confusion to Confusion: Issues and the American Voter (1956–72)." *American Political Science Review* 71:31–43.

Nie, N. H., and K. Anderson. 1974. "Mass Belief Systems Revisited: Political Change and Attitude Structure." *Journal of Politics* 36:540–91.

Pierce, J. C. 1970. "Party Identification and the Changing Role of Ideology in American Politics." *Midwest Journal of Political Science* 14:25–42.

Pierce, J. C., and D. D. Rose. 1974. "Non-attitudes and American Public Opinion." *American Political Science Review* 68:626–49.

Pomper, G. M. 1972. "From Confusion to Clarity: Issues and American Voters." *American Political Science Review* 66:415–28.

Rosenberg, M. 1968. *The Logic of Survey Analysis*. New York: Basic.

Schuman, H., and S. Presser. 1981. *Questions and Answers in Attitude Surveys*. New York: Academic Press.

Stimson, J. A. 1975. "Belief Systems: Constraint, Complexity, and the 1972 Election." *American Journal of Political Science* 19:393–417.

Sullivan, J. L., J. E. Piereson, and G. E. Marcus. 1978. "Ideological Constraint in the Mass Public: A Methodological Critique and Some New Findings." *American Journal of Political Science* 22:233–49.

Wiley, D. E., and J. A. Wiley. 1970. "The Estimation of Measurement Error in Panel Data." *American Sociological Review* 35:112–17.

A Note on Emotionality, Self, and Interaction[1]

Norman K. Denzin
University of Illinois, Urbana-Champaign

Recent work in the sociology of emotions is reviewed. The phys-
iological bias in that work is noted, as is the emphasis on the study
of emotion as either the cause or the consequence of factors external
to emotional consciousness and interaction. Emotions are reconcep-
tualized as self-feelings. Their study from a symbolic interactionist
point of view is suggested.

After decades of writing superficially about the emotions, sociologists are
now taking them seriously and attempting to construct theories which
would give human emotionality a central place in the sociological picture
of persons and their societies (Scheff 1979, 1983). Recent articles in this
journal (Shott [1979], Hochschild [1979], and Kemper [1981]) attempting
to contribute to a sociology and social psychology of the emotions are to
be applauded. Shott's application of principles from symbolic interac-
tionism, Hochschild's conceptions of feeling rules and emotion manage-
ment strategies, and Kemper's social relational theory of emotion represent
significant advances over previous formulations. This body of work has
several distinct advantages. First, it treats emotion sociologically, rela-
tionally, and interactionally. Second, it synthesizes wide-ranging literature
in experimental psychology, physiology, social psychology, anthropology,
and sociology. Third, it offers innumerable propositions which might be
explained empirically and comparatively in different societies, cultures,
and social classes. Fourth, it speaks to the emotions as sociological, not
strictly physiological or psychological, processes.

These advantages notwithstanding, this body of work suffers from two
basic problems which need correcting. The first is definitional and in-
volves the way in which the phenomena of emotion are to be formulated
sociologically. The second problem is one of domain priorities and involves
the place of physiological or somatic processes in a sociological theory of
emotion. The two problems tend to blur.

THE DEFINITION OF EMOTION

Hochschild, Shott, and Kemper treat emotions as states of physiological arousal. Shott (1979, p. 1318) states: "I shall use Schachter's (1971, pp. 23–24) conception of emotion as a state of physiological arousal defined by the actor as emotionally induced." Hochschild (1979, p. 551, n. 2) defines emotion as "bodily cooperation with an image, a thought, a memory—a cooperation of which the individual is aware." Kemper (1981, p. 339) states that "different outcomes in power and status relations instigate different physiological processes, which are in turn related to different emotions."

CRITICISM

The following diagram depicts the causal interpretation of the emotion proposed by these authors.

Bodily ——————→ overt   ←————————→ cultural factors   ——————→ emotional
reactions   behaviors   relational factors   labeling
                                           feeling rules
                                           structural factors
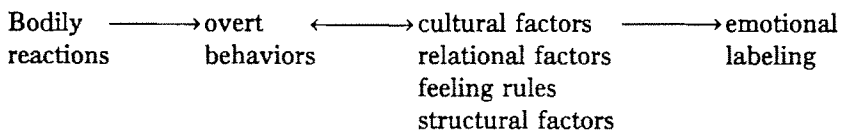
The causal model is clear. Bodily changes produce overt behaviors, which are mediated and defined by cultural, relational, structural, and feeling rule factors. These factors, in turn, produce variations in the labeling of emotional states. Emotion is seen as either the cause or the consequence of factors external to emotional consciousness. Emotions are treated like stones or ponds or static objects which are given labels (Sartre [1939] 1962).

Emotions are embodied experiences. The place of the body as an instrument in the expression of emotion cannot be denied. However, the body does not call out its own interpretations of internal somatic states. These interpretations are given in the field of consciousness by the person, reflectively, through self-interactions. Shott, Kemper, and Hochschild ignore the place of self-interactions and self-referencing in their definitions of emotion.[2]

Pure behavior is not emotion, nor is the simple awareness of that behavior emotion (Sartre [1939] 1962, p. 74). Behavior becomes emotional only when it is so interpreted by the person and brought into self-inter-

---

[2] Hochschild does (1979, pp. 554–55, 558) call for a theory of self and the emotions. Her notion of the self as emotion manager is suggestive. Emotions, though, are processes, not things. Furthermore, what is being managed is not emotion, but the self-in-the-feeling that is felt.

actions. The body adjusts itself to the emotional interpretations persons place upon it. The bodily disturbances which occur during emotional experiences ratify for persons the reality of the experience they are feeling. Bodily disturbances provide the belief in an emotion as it is lived (Sartre [1939] 1962, pp. 78–79). Such disturbances do not in and of themselves cause the emotion that is felt. This, of course, reverses James's ([1890] 1950) famous definition of emotion. His definition was physiological; mine is interactional.

Emotion is a social, interactional, linguistic, and physiological process that draws its resources from the human body, from human consciousness, and from the world that surrounds a person. The sociological analysis of emotion cannot begin, however, with the physiological and neurological structures of the human body and the brain. Nor can such a study begin with the examination of factors that lie outside the interaction process wherein emotion is experienced. Social, cultural, relational, and rule-embodied interpretations of emotion must be temporarily suspended, while emotion is defined and studied from within, as a lived, interactional process.

EMOTION REDEFINED

Emotion may be succinctly defined as *self-feeling*. More specifically, *emotions are temporally embodied self-feelings which arise from emotional social acts persons direct to self or have directed toward them by others.* Emotions are lodged in social acts and self-interactions. It is my thesis that all the emotional terms used in everyday language, including being angry, resentful, sad, fearful, joyful, depressed, hostile, enraged, ashamed, proud, affectionate, friendly, embarrassed, rejected, guilty, or in pain, refer to embodied feelings, mental states, interactional experiences, and judgments of others (real and imagined) that persons feel and direct (or have directed) toward self. See James ([1890] 1950), 1:305–6; 2:442–86; Cooley (1902, p. 184); Smith (1892, p. 164); Bain (1859, p. 125); Scheler ([1913] 1970, p. 10; 1973, p. 259); Heidegger (1982, p. 137); Sartre ([1939] 1962, p. 29).

Self-feelings refer to *any emotion* a person feels, including bodily sensations, sensible feelings, intentional feeling-states, and feelings of self as a moral, sacred, or profane object (Scheler 1973, pp. 330–31). The feelings that a person feels have a threefold structure: (1) a sense of the feeling in terms of awareness and definition; (2) a sense of the self feeling the feeling; (3) a revealing of the moral or feeling self through this experience (Heidegger 1982, p. 137). The feeling person, the person in emotional consciousness, feels his or her self in emotion. An emotion that does not in

some way have the self or the self-system of the person as its referent seems inconceivable.

STUDYING EMOTION AS SYMBOLIC INTERACTION

Defining emotions as self-feelings has the advantage of placing their study directly in the field of social interaction. Mead's (1938) and Blumer's (1969) views of the self, the social act, symbolic interaction, significant gestures, the I and the me, self-interaction, the triadic structure of meaning, social objects, and joint acts can now be brought directly to bear upon the study of emotion.[3]

Self-interaction

If emotions are conceptualized as self-feelings, they can be seen as arising out of the self-interactions persons direct toward themselves and from the reflected appraisals of others (imagined and real). Figure 1 describes the process of emotional self-interaction (see Denzin 1980, p. 255).
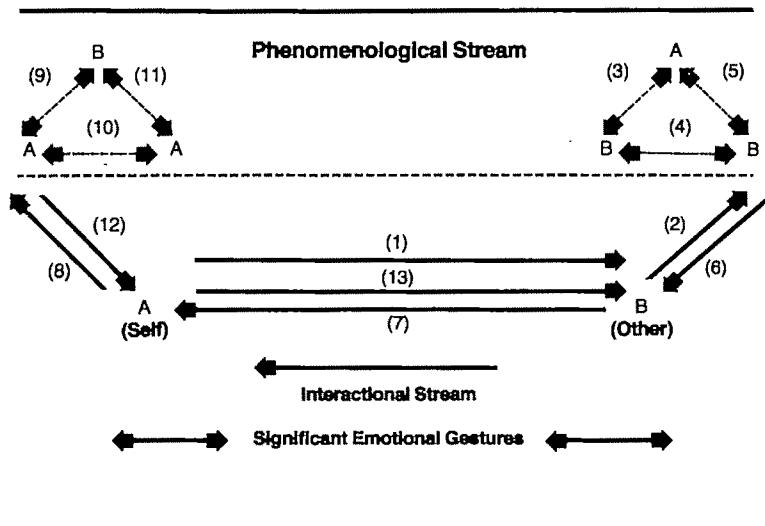
FIG. 1.—Emotional self-interaction in the social situation. Numbers 1–13 suggest the temporal flow of emotional self-interaction, starting with A's first act toward B.

[3] This research note may be read as "G. H. Mead's Interactionist Theory of Emotions." Mead's (1895, 1982) few remarks on emotion were extensions and elaborations of James's and Dewey's (1894, 1895) physiological positions. The present interpretation is developed in Denzin (1983).

The *phenomenological stream* describes the inner side of interaction that occurs when the person interacts with himself and with another in a social situation. In the phenomenological stream the person takes his own attitude toward self, (A–A), toward the other, (A–B), and turns the imagined attitude of the other toward himself, (B–A). The *interactional stream* references the co-present, face-to-face interaction that occurs between two parties in a concrete or an imagined situation. *Emotional self-interaction* involves person A initiating a line of action toward B that calls out in B a significant emotional gesture that is present in A's action and thought. Person A turns this emotional gesture inward, judging and interpreting the meaning of B's actions in light of A's own incipient emotional attitude(s) (A–A, A–B, B–A). This interpretation becomes part of A's emotional self-feelings and is then incorporated into A's next gesture or statement to B. Thus, A's self-feeling becomes part of an emotional social act that enters B's inner phenomenological stream and becomes part of B's emotional social act (toward both B and A).

Social Acts

Emotions may be studied as social acts, with covert and overt, inner and outer dimensions. Emotions have phases, including beginnings, middles, and ends. Many emotional acts are retrospective, backward looking, locked in the temporality of the past. Others lie in anticipated futures and some are realized in the immediate present (Heidegger 1962). Emotional acts may be blunted, checked, or redefined. They may be reflectively and unreflectively organized (Athens 1980).

Temporality

The temporality of emotion is circular (Denzin 1982). The future, the past, and the present are vividly interconnected in the emotional acts of a person. The temporality of emotion as lived experience blurs the distinctions between the past, the present, and the future. Intense emotionality appears to stop time.

Others

Emotions without at least the previous presence of others are unimaginable.[4] Emotional associates are central to the study of the social acts that produce and make up emotion. If all emotional terms reference feelings

[4] These others may be termed *emotional associates.* An emotional associate is any person who is drawn, in some fashion, into the emotional world of experience of another individual.

directed toward the self, they also reference feelings the person feels toward others, including social objects. Hence, emotional terms are interactive referents. Emotional terms link selves together in joint activities or social acts (see fig. 1). Others enter the inner side of emotionality through significant symbols and *"other-taking"* emotions (Shott 1979, pp. 1325–30).

Emotional Reality

Emotional acts have a lived "realness" that is not doubted. The emotional experience, in the form of embodied self-feelings, radiates through the inner and outer streams of experience and moves persons into the world of emotion. This is a world that is neither wholly rational nor fully deterministic, nor is it a world where things are entirely as they appear to be. It is a world filled with nuance, symbolism, and hidden, indirect meaning.

The Circuit of Selfness

A circuit of selfness (Sartre 1956, pp. 155–58) attaches a person to the world. In that circuit, emotionality, meaning, and the others are joined. This field of experience anchors the self in the other's phenomenological and interactional streams. It is in the merger of these two streams of experience that emotional self-feelings are lodged.

There are deep and surface levels of the person which are reflected in deep and surface self-feelings and emotions. Surface emotions reflect the taken-for-granted feelings of the person. Deep emotions reflect core, inner feelings of self. Guilt, shame, and resentment are deep emotions. Deep feelings reflect the feelings of the inner moral person (Heidegger 1982). Both levels of self-feeling are lodged in the circuit of selfness that attaches the person to the world. This *circuit of self-feeling* describes the interrelationships between the two levels of the self.

There is a double structure to this feeling experience. In any emotional self-feeling there is a *core* or essential emotional experience that is named and felt—perhaps anger, or fear, or guilt. Surrounding that core feeling is a *horizon* of interrelated feelings, thoughts, and memories which are passed through, as in *emotional memory,* when the person moves into a core emotional experience (Stanislavski 1936, p. 158). This is called the *horizon of emotion.*

CONCLUSIONS

Emotions are not mere cognitive responses to physiological, cultural, or structural factors. They are interactive processes best studied as social

acts involving self and other interactions. The sociological study and definition of emotion must begin with the study of selves and others, joined and separated in episodes of co-present interaction. Defining emotion as self-feeling returns the sociology of emotion to the world of lived, interactional experience. The foregoing remarks are intended to suggest some directions such an inquiry might take.[5]

REFERENCES

Athens, Lonnie H. 1980. *Violent Criminal Acts and Actors*. Boston: Routledge & Kegan Paul.
Bain, Alexander. 1859. *The Emotions and the Will*. London: Parker.
Blumer, Herbert. 1969. *Symbolic Interactionism*. Englewood Cliffs, N.J.: Prentice-Hall.
Cooley, Charles H. 1902. *Human Nature and the Social Order*. New York: Scribner.
Denzin, Norman K. 1980. "A Phenomenology of Emotion and Deviance." *Zeitschrift für Soziologie* 9 (July): 252–61.
———. 1982. "The Temporality of Everyday Moods." Presented to the 1982 annual meetings of the Pacific Sociological Society, San Diego, April 22.
———. 1983. *On Understanding Emotion*. San Francisco: Jossey-Bass.
Dewey, John. 1894. "The Theory of Emotion. I. Emotional Attitudes." *Psychological Review* 1 (November): 553–69.
———. 1895. "The Theory of Emotion. II. The Significance of Emotions." *Psychological Review* 2 (January): 13–32.
Heidegger, Martin. 1962. *Being and Time*. New York: Harper & Row.
———. 1982. *The Basic Problems of Phenomenology*. Bloomington: Indiana University Press.
Hochschild, Arlie. 1979. "Emotion Work, Feeling Rules, and Social Structure." *American Journal of Sociology* 85 (November): 551–75.
James, William. (1890) 1950. *The Principles of Psychology*. New York: Dover.
Kemper, Theodore D. 1981. "Social Constructionist and Positivist Approaches to the Sociology of Emotions." *American Journal of Sociology* 87 (September): 336–62.
Mead, George Herbert. 1895. "A Theory of Emotions from the Physiological Standpoint." *Psychological Review* 2 (January): 162–64.
———. 1938. *Philosophy of the Act*. Chicago: University of Chicago Press.
———. 1982. *The Individual and the Social Self: Unpublished Work of George Herbert Mead*, edited with an introduction by David L. Miller. Chicago: University of Chicago Press.
Sartre, Jean-Paul. (1939) 1962. *Sketch for a Theory of the Emotions*. Translated by Philip Mariet with a preface by Mary Warnock. London: Methuen.
———. 1956. *Being and Nothingness*. Translated by Hazel E. Barnes. New York: Philosophical Library.
Schachter, Stanley S. 1971. *Emotion, Obesity and Crime*. New York: Academic Press.
Scheff, Thomas. 1979. *Catharsis in Healing, Ritual and Drama*. Los Angeles and Berkeley: University of California Press.
———. 1983. "Current Issues in the Social Psychology of Emotions. *Annual Review of Sociology* 9, in press.
Scheler, Max. 1913. *The Nature of Sympathy*. Translated by Peter Heath with an Introduction by W. Stark. Hamden, Conn.: Archon.
———. 1973. *Formalism in Ethics and Non-Formal Ethics of Value: A New Attempt toward*

---

[5] I have suggested some of these directions in Denzin (1983), including an examination of the structures and stratification of lived emotion, the problem of emotional intersubjectivity, and the study of enacted emotion.

the Foundation of an Ethical Personalism. Translated by Manfred S. Frings' and Roger L. Funk. Evanston, Ill.: Northwestern University Press.

Shott, Susan. 1979. "Emotion and Social Life: A Symbolic Interactionist Analysis." *American Journal of Sociology* 84 (May): 1317–34.

Smith, Adam. 1892. *The Theory of Moral Sentiments*. London and New York: Bell.

Stanislavski, Constantin. 1936. *An Actor Prepares*. New York: Theatre Arts Books.

Status in the World System, Income Inequality, and Economic Growth[1]

Patrick D. Nolan
University of South Carolina

This paper addresses recent challenges to fundamental world-system research findings by using methods and measures suggested by critics of these findings. Reanalysis of income and economic growth data using a categorical indicator of world-system status derived from Snyder and Kick suggests that dependency does increase income inequality and retard per capita economic growth net of initial development. This offers renewed support for world-system theory by rebutting Weede's criticism of Rubinson, and Jackman's criticism of Snyder and Kick.

INTRODUCTION

Two recent challenges have been offered to world-system research. First, findings that dependency increases income inequality have been criticized for their failure to control for the curvilinear effects of development (Weede 1980). Second, the observed effects of dependency on rates of per capita economic growth have been claimed to be artifacts of inadequate measures of growth and resulting heteroscedasticity (Jackman 1980). In this analysis I employ procedures and measures suggested by these and other critics to provide a more adequate test of the original hypotheses. (See Appendix for cases used in this reanalysis.)

Rubinson (1976) and Rubinson and Quinlan (1977) presented evidence that there was significantly greater income inequality in dependent than in core nations. They found that a much greater share of income was concentrated in the upper regions of the income distributions of dependent

countries, while in core countries, although the upper regions continued
to have the largest shares, proportionately more income accrued to the
middle regions. This was argued to be evidence of a small labor aristocracy
in the periphery and a relatively strong middle class in the core, as pre-
dicted in world-system/dependency theory.

Weede (1980), however, has challenged these findings by questioning
Rubinson's control for the effects of development on income distribution.
According to Weede, a second-order polynomial is necessary, and Ru-
binson's use of the log of development is inadequate as a control because
economists have long maintained that development has a curvilinear
nonmonotonic effect on inequality. At low levels of development, growth
may actually increase inequality, whereas growth at higher levels reduces
it. To demonstrate that dependency has an effect on inequality indepen-
dent of the simple effects of development, therefore, it is necessary to use
a control that will detect this expected reversal in the direction of the
relationship. However, when Weede (1980) attempted to replicate Ru-
binson's (1976) analysis employing polynomial controls, all the regression
coefficients for Rubinson's indicators of dependency failed to reach sig-
nificance at even the .10 level. From this, Weede concluded that Ru-
binson's results were due to his "misspecification" of the economic model
and rejected the hypothesis that position in the world economy affects
economic inequality net of economic development.

To address this conclusion, I reanalyzed Paukert's (1973) and the World
Bank's (1980) data on income inequality, using a network-based measure
of status in the world economy derived from Snyder and Kick (1979),
which they argue is conceptually superior to the measures of dependency
used by Rubinson and others, and employing Weede's preferred poly-
nomial control for development, the log of energy consumption per capita,
and its square.[2] The results of this reanalysis are presented in table 1.
The first column reports the results of regressing income inequality and
income share on world-system status without controls for development.
Clearly these zero-order results resemble Rubinson's.[3] Although it is only
for the upper-middle quintile (61%–80%) that both the semiperiphery and
the periphery are significantly different from the core (values for the core,
as the omitted reference category, are indicated by the intercept), the

[2] In order to detect the effects of "status level," or hierarchical position, in the world political
economy, the nine "blocks" found by Snyder and Kick (1979) have been aggregated into
three categories on the basis of their correspondence with core, semiperiphery, and periphery
"roles" (see Snyder and Kick 1979, pp. 1110–16 and below). For evidence of the construct
validity (Selltiz et al. 1959, pp. 159–63) of the aggregated measure, see Nolan (1981, 1982).

[3] This replication is made even more interesting by the fact that the status measure is not
related significantly to Rubinson's primary indicators of dependency: state strength, gov-
ernment revenue as a percentage of GNP ($F = 2.166$, $P = .126$); and trade dependence,
trade as a percentage of GNP ($F = .603$, $P = .551$).

411

TABLE 1

HOUSEHOLD INCOME REGRESSED ON STATUS AND DEVELOPMENT POLYNOMIAL

	Zero-Order Semiperiphery and Periphery Contrasts with Core	Semiperiphery and Periphery Contrasts with Core Controlling Development	Periphery Contrasts with Core and Semiperiphery Controlling Development
	Paukert Data Circa 1965 ($N=48$)		
Gini index of inequality:			
Intercept	.398***	−.121 N.S.	−.120 N.S.
Ldev	. . .	.431***	.428***
Ldev²	. . .	−.080**	−.080***
Semiperiphery	.032 N.S.	−.003 N.S.	. . .
Periphery.	.092**	.074 N.S.	.076**
R^2 change	. . .	.108*	.108**
Multiple R^2	.171**	.369***	.369***
Equalization percentage:			
Intercept	29.538***	−15.838 N.S.	−14.795 N.S.
Ldev	. . .	38.661***	37.501***
Ldev²	. . .	−7.358***	−7.072***
Semiperiphery	2.831 N.S.	−1.020 N.S.	. . .
Periphery.	8.598**	5.937 N.S.	6.746**
R^2 change	. . .	.119*	.118**
Multiple R^2	.209**	.420***	.418***
81–100 percentile:			
Intercept	46.762***	−3.639 N.S.	−3.343 N.S.
Ldev	. . .	45.800***	44.381***
Ldev²	. . .	−8.911***	−8.561***
Semiperiphery	4.277 N.S.	−1.247 N.S.	. . .
Periphery.	10.648***	6.184 N.S.	7.173**
R^2 change	. . .	.107*	.106*
Multiple R^2	.245**	.473***	.471***
61–80 percentile:			
Intercept	22.608***	37.245***	37.182***
Ldev	. . .	−16.087***	−15.785***
Ldev²	. . .	3.378***	3.303***
Semiperiphery	−2.900*	.266 N.S.	. . .
Periphery.	−4.662***	−1.420 N.S.	1.881 N.S.
R^2 change	. . .	.035 N.S.	.035 N.S.
Multiple R^2	.281***	.481***	.480***
41–60 percentile:			
Intercept	15.715***	28.140***	27.898***
Ldev	. . .	−13.517***	−12.359***
Ldev²	. . .	2.828***	2.543***
Semiperiphery	−1.592 N.S.	1.018 N.S.	. . .
Periphery.	−4.274***	−1.623 N.S.	−2.430**
R^2 change	. . .	.104*	.096**
Multiple R^2	.318***	.493***	.485***

NOTE.—Ldev and Ldev² = logarithm of 1965 energy consumption per capita in kilograms coal equivalent (from Taylor and Hudson 1973, p. 84 [V157]) and its square. Semiperiphery and periphery = status in the world economy (from Snyder and Kick 1979); see Appendix. N.S. = nonsignificant.

* Significant at .05.
** Significant at .01.
*** Significant at .001.

TABLE 1 *(Continued)*

	Zero-Order Semiperiphery and Periphery Contrasts with Core	Semiperiphery and Periphery Contrasts with Core Controlling Development	Periphery Contrasts with Core and Semiperiphery Controlling Development
	Paukert Data Circa 1965 ($N = 48$)		
21–40 percentile:			
Intercept	10.131***	22.885***	22.836***
Ldev	. . .	−10.575**	−10.341**
Ldev^2	. . .	1.967**	1.910**
Semiperiphery	−.623 N.S.	.206 N.S.	. . .
Periphery	−1.990*	−1.568 N.S.	−1.731*
R^2 change	. . .	.081 N.S.	.081*
Multiple R^2	.118 N.S.	.292*	.291**
0–20 percentile:			
Intercept	4.931***	15.630***	15.747***
Ldev	. . .	−5.680*	−6.241*
Ldev^2	. . .	.746 N.S.	.884 N.S.
Semiperiphery	.600 N.S.	−.494 N.S.	. . .
Periphery	.124 N.S.	−1.746 N.S.	−1.354*
R^2 change	. . .	.077 N.S.	.073*
Multiple R^2	.013 N.S.	.302**	.297***
	World Bank Data 1970		
96–100 percentile ($N = 38$):			
Intercept	15.900***	.335 N.S.	2.156 N.S.
Ldev	. . .	22.618 N.S.	19.092 N.S.
Ldev^2	. . .	−4.976 N.S.	−4.193 N.S.
Semiperiphery	5.922 N.S.	−2.361 N.S.	. . .
Periphery	14.325***	5.827 N.S.	7.712*
R^2 change	. . .	.100 N.S.	.097*
Multiple R^2	.329***	.385**	.382***
0–20 percentile ($N = 54$):			
Intercept	5.923***	12.584**	13.253**
Ldev	. . .	−3.754 N.S.	−5.158 N.S.
Ldev^2	. . .	.525 N.S.	.841 N.S.
Semiperiphery	−.464 N.S.	−.941 N.S.	. . .
Periphery	−1.690**	−2.356***	−1.562**
R^2 change	. . .	.160**	.143*
Multiple R^2	.163**	.230**	.212**

Note.—Ldev and Ldev^2 = logarithm of 1965 energy consumption per capita in kilograms coal equivalent (from Taylor and Hudson 1973, p. 84 [V157]) and its square. Semiperiphery and periphery = status in the world economy (from Snyder and Kick 1979); see Appendix. N.S. = nonsignificant.

* Significant at .05.
** Significant at .01.
*** Significant at .001.

413

statuses are appropriately rank ordered for both summary measures and in four of the five income quintiles.[4] The lowest quintile (0%–20%) constitutes the exception. The status dummies also explain a significant proportion of variance in both summary measures and in three of the five quintiles (41%–60%, 61%–80%, 81%–100%). Similar patterns are found with data on the tails of income distributions reported by the World Bank (1980) presented in the bottom panels of table 1.

The increase in explained variance that accompanies introduction of a polynomial control for development, which can be seen by comparing the first and second columns, supports Weede's "specification of development" argument, but use of it as a control does not eliminate the effects of status on inequality. When entered after the polynomial development control, the status dummies continue to produce a significant increment in explained variance in both summary measures, 10.8%, 11.9%; and in two income quintiles (41–60, and 81–100 percentiles), 10.4%, 10.7%. The major effect of the control is to reduce further, and nominally reverse, the contrast of semiperiphery and the core without substantially affecting the differences between the core and periphery. This suggests that whatever differences exist between core and semiperiphery are due primarily to differences in their levels of development and do not reflect any fundamental consequence of their differing world-economic roles. Since this is the case, it makes sense to combine the two into a single reference category with which the periphery can be compared, net of the effects of development. Examination of the third column indicates that when this is done the effects of periphery status emerge even more strongly. In four of the five quintiles, and for both summary measures, there is a significant increase in explained variance. Ironically, it is only in the fourth quintile (61%–80%), where both status coefficients were significant before controls, that the increase in explained variance is nonsignificant. Similar results are also found using the more restricted World Bank (1980) data.

Thus it is evident that, although controls do reduce the already smaller semiperiphery/core differences, they do not alter significantly the contrasts of the periphery and the core. It must be concluded, therefore, that the effect of status, or dependency, on income inequality is not simply an artifact of failure to control adequately for economic development. When

---

[4] Since cases in this analysis do not constitute a probability sample because all nations for which comparable data were available were included, significance values do not have their conventional meaning. In fact, the population character of the data led Snyder and Kick (1979) to deemphasize the significance and stress the size of their unstandardized regression coefficients in their analysis. Jackman (1980) and Weede (1980), however, relied much more on significance values in making decisions about the importance of relationships. Without taking a position on the general issue, the present analysis examines significance values because they have been used by these critics to challenge some of the research reexamined here.

the network-derived measure of status is employed, position is found to have significant and substantial effects on income distribution, net of the curvilinear effects of simple development. Fundamental differences in income distribution and inequality between the core and periphery remain when development is controlled.

The relationship between world economic position and rate of economic growth is perhaps even more controversial than that between position and inequality. Most recently, Jackman (1980) has questioned the findings of Snyder and Kick (1979) and others (e.g., Chase-Dunn 1975) that, in recent decades, noncore nations have experienced lower rates of economic growth than core nations. Jackman argues these results may be an artifact of the measure of economic growth used by these researchers, GNP per capita differences at two time points, the use of which also produces heteroscedastic disturbance in their regression analyses. According to Jackman, both of these interrelated problems can be avoided by using a continuous measure of economic growth rate. However, when he attempted to replicate Snyder and Kick's analysis using such a measure, he found that, despite the fact that all nine noncore blocks grew more slowly than the core, not a single regression coefficient was "significant." He concluded, therefore, that no convincing evidence of a relationship between world economic position and rate of economic growth had yet been offered.

Jackman's (1980) argument for the superiority of a continuous over a simple difference measure of growth is well founded, but when he continued to use all nine noncore blocks as regressors in his reanalysis, he perpetuated another theoretical weakness in Snyder and Kick's original analysis. Use of the nine blocks that *constitute* the statuses may, in fact, obscure the effect of the statuses. Since the theory maintains, and Snyder and Kick claim to have confirmed, that there are three statuses in the world system, it is important and necessary to compare the *statuses* with one another, rather than compare the smaller clusters of nations that, together, constitute them. To test the hypothesis properly, therefore, it is necessary not only to take Jackman's advice and use the continuous measure of growth (cf. World Bank 1980, p. 5, n. 2), and to focus on the time period he considers more reliable statistically (1960–70), but, in addition, to abandon the false precision of entering all nine subgroups as predictors and, instead, use the status levels they constitute. It is also important to consider the possibility that, as was the case with inequality, development may be curvilinearly (and nonmonotonically) related to growth.

The first column of table 2 displays the results of regressing average annual rate of economic growth, GNP per capita, and total GNP with population as a predictor, on controls for school enrollment, initial de-

velopment, the log of energy consumed per capita, and status. There is
a significant increment to variance explained when the status dummies
are entered after controls (8.1%), but only the regression coefficients for
the periphery are significant. However, the fact that initial development
has a nonsignificant effect on growth rate, as in Jackman's reanalysis,
leads to the suspicion that the log of development is not sufficient to detect
the relationship between level of development and rate of economic growth.
A simple way to test this possibility is to enter power terms of development
hierarchically and determine if they increase the explained variance sig-
nificantly (Cohen and Cohen 1975; Jagodzinski and Weede 1981). When
this is done, both the square and the cube of development can be seen
to have significant effects, indicating that a cubic function best describes
the relationship between initial development and the rate of economic
growth. The most *conservative* test of the effect of status, therefore, would
be provided by entering the status dummies only after school enrollment

TABLE 2

AVERAGE ANNUAL RATE OF ECONOMIC GROWTH REGRESSED ON INITIAL
DEVELOPMENT, SCHOOL ENROLLMENT, AND POPULATION
GROWTH RATE 1960–70 ($N = 95$)

	Controlling Simple Function of Development	Controlling Quadratic Function of Development	Controlling Cubic Function of Development
GNP per capita growth rate:			
Intercept	2.938*	−.003 N.S.	5.748*
Ldev	−.004 N.S.	3.933**	−5.730 N.S.
Ldev²	. . .	−.927**	3.901 N.S.
Ldev³	. . .	. . .	−.718*
School enrollment 1960	.019 N.S.	.022 N.S.	.018 N.S.
Semiperiphery	−.404 N.S.	−1.425*	−1.656**
Periphery.	−1.775**	−2.685***	−2.751***
R² change status	.081**	.123***	.123***
Multiple R²	.320**	.396***	.432***
Adjusted R²	.290**	.362***	.393***
GNP growth rate:			
Intercept	2.859*	−.034 N.S.	5.732*
Ldev	.003 N.S.	4.049**	−5.788 N.S.
Ldev²	. . .	−.947**	3.948 N.S.
Ldev³	. . .	. . .	−.727*
School enrollment 1960	.019 N.S.	.021 N.S.	.018 N.S.
Population growth rate			
1960–70	1.061***	.965***	1.037***
Semiperiphery	−.426 N.S.	−1.350 N.S.	−1.663*
Periphery.	−1.831***	−2.640***	−2.788***
R² change status	.081**	.111***	.114***
Multiple R²	.278***	.356***	.394***
Adjusted R²	.238***	.314***	.346***

NOTE.—Ldev = log of energy consumption per capita in kilograms coal equivalent. Adjusted school enrollment 1960 and energy consumption 1960 data from Taylor and Hudson (1973, p. 63 [V109], p. 83 [V156]); GNP growth rate from the World Bank (1980); GNP per capita growth rate = 1 + GNP GR/1 + POP GR. Ldev³ significantly increases R² in all equations (P < .05). N.S. = nonsignificant.
 * Significant at .05.
 ** Significant at .01.
 *** Significant at .001.

and the (quadratic and) cubic functions of development have been allowed to explain as much variance as they can.

When this is done, instead of attenuating, the effects of status actually increase. Both regression coefficients are significant and appropriately rank ordered, and the status dummies together explain 12% of the variance in per capita rates and a little over 11% in the total growth rates. This is so despite the fact that the polynomial of development has a significant and sizable effect on growth rates. The originally nonsignificant effect of development appears to have been the result of attempting to fit a monotonic function to a nonmonotonic relationship. The better fitting development function does not "explain away" the effect of status, but it does confirm Jagodzinski's and Weede's (1981) suspicion that many important sociological relationships are not simple linear or curvilinear functions.

DISCUSSION AND CONCLUSIONS

Reexamination of recently controversial world-system research findings using improved measures and models has provided support for the original world-system/dependency predictions. A substantial relationship is found between income inequality and position in the world system despite a more rigorous polynomial control for development. Reanalysis of the effects of position on relative rates of economic growth has also provided reconfirmation of Snyder and Kick (1979) and rebutted Jackman's (1980) challenge. Although the criticisms were well founded in both cases, the network-based measure of status revealed a continuing effect after more reliable measures and more rigorous specifications of controls were used. Certainly much more research is needed to test these theoretical predictions fully, but the present analysis indicates that it would be premature to dismiss them on the basis of existing criticism. It also clearly indicates that it is important to examine the possibility of nonlinearity and nonmonotonicity before specifying controls for basic structural and developmental features of nations in cross-national research.

APPENDIX

Listing of Nations by Status (after Snyder and Kick 1979, p. 1110)

Blocks are designated by the first capital letter in each paragraph. If cases are included in only some of the analyses, letters indicate which analyses they are in: E = economic growth, P = Paukert, B = bottom 20% World Bank, T = top 5% World Bank.
Core
C: Canada, United States, United Kingdom, Netherlands, Belgium (E),

Luxembourg (E), France, Switzerland (E), Spain (E), Portugal (B), West Germany (E, B), Austria (E), Italy (E, P, B), Yugoslavia (E, B, T), Greece (E, P), Sweden, Norway, Denmark, South Africa (E, P), Japan, Australia.

Semiperiphery

C': Venezuela (E, B, P), Peru (E, B, P), Argentina (E, B, P), Uruguay (E, B), South Korea.

D: Ireland (E), Cyprus (E, B, T), Kenya (E, B, T), Iran (E, B, T), Turkey (E, B, T), Iraq (E, P), Lebanon (P), Israel.

D': Finland (E, P), Saudi Arabia (E), Taiwan (E, B), India, Pakistan, Burma (E, B, T), Ceylon (E, B, P), Malaysia (E, B, T), Philippines (E, B, P).

Periphery

A: Chad (E, P), Congo (Brazzaville) (E), Congo (Kinshasa) (E), Uganda (E, B, T), Burundi (E), Ruanda (E), Somalia (E), Ethiopia (E), Morocco, Algeria (E), Tunisia (E, P), Sudan (E, B, P), United Arab Republic (E, B, T).

B: Mali (E, B, T), Mauritania (E), Ghana (E), Upper-Volta (E), Senegal (E, P), Dahomey (P), Niger (E, P), Ivory Coast (E, P), Republic of Guinea (E), Liberia (E, B, T), Sierra Leone, Togo (E), Cameroon (E), Nigeria (E, P), Gabon, Central African Republic (E).

E: Panama, Colombia, Ecuador (E, P), Brazil (E, B, P), Bolivia, Paraguay (E, B, T), Chile (E, B, P).

E': Haiti (E), Dominican Republic (E, B, T), Mexico, Guatemala (E, B, T), Honduras (E, B), El Salvador (E, P), Nicaragua (E, B, T), Costa Rica.

F: Jamaica (E, P), Trinidad and Tobago, Malta (E), Nepal (E), Thailand (E, B, T), New Zealand (E, B, T), Iceland (E).

F': Syria (E), Afghanistan (E), Indonesia (E, B, T).

REFERENCES

Chase-Dunn, Christopher. 1975. "The Effects of International Economic Dependence on Development and Inequality: A Cross-national Analysis." *American Sociological Review* 41:660–75.
Cohen, Jacob, and Patricia Cohen. 1975. *Applied Regression/Correlation Analysis for the Behavioral Sciences.* New York: Wiley.
Jackman, Robert. 1980. "A Note on Measurement of Growth Rates in Cross-national Research." *American Journal of Sociology* 86:604–17.
Jagodzinski, Wolfgang, and Erich Weede. 1981. "Testing Curvilinear Propositions by Polynomial Regression with Particular Reference to the Interpretation of Standardized Solutions." *Quality and Quantity* 15:447–63.
Nolan, Patrick D. 1981. "Status in the World Economy and the Socioeconomic Structure of Nations." Paper presented at the 76th annual meeting of the American Sociological Association, Toronto, Ontario.
———. 1982. "Stratification in the World System." *Current Anthropology* 22:193–94.
Paukert, Félix. 1973. "Income Distribution of Different Levels of Development: A Survey of Evidence." *International Labour Review* 108:97–125.

Rubinson, Richard. 1976. "The World-Economy and the Distribution of Income within States: A Cross-national Study." *American Sociological Review* 41:638–59.

Rubinson, Richard, and Dan Quinlan. 1977. "Democracy and Social Inequality: A Reanalysis." *American Sociological Review* 42:611–22.

Selltiz, Claire, Maria Jahoda, Morton Deutsch, and Stuart W. Cook. 1959. *Research Methods in Social Relations*. New York: Holt, Rinehart & Winston.

Snyder, David, and Edward Kick. 1979. "Structural Position in the World System and Economic Growth 1955–70: A Multiple Network Analysis of Transnational Interactions." *American Journal of Sociology* 84:1096–1126.

Taylor, Charles, and Michael Hudson. 1973. *World Handbook of Political and Social Indicators*. 2d ICPR ed. Ann Arbor: Inter-University Consortium for Political Research.

Weede, Erich. 1980. "Beyond Misspecification in Sociological Analyses of Income Inequality." *American Sociological Review* 45:497–501.

World Bank. 1980. *World Tables*. 2d ed. Baltimore: Johns Hopkins University Press.

Response Variability in Self- and Proxy Reports of Paternal and Filial Socioeconomic Characteristics[1]

Michael P. Massagli and Robert M. Hauser
University of Wisconsin—Madison

In a 1973 Australian survey, Broom et al. located two distinct sub-samples of men in which fathers and sons reported their own status characteristics and those of their sons and fathers, respectively. Inconsistencies between these two subsamples led Broom et al. to infer that "respondents report (perhaps unconsciously) their socio-economic careers and the careers of their kin in such ways as either to maximize differences between themselves and their fathers or to minimize differences between themselves and their sons." If true, this would invalidate most stratification research. Reexamination of the Australian data suggests that the paternal and filial reports are not directly comparable. In a sample of Wisconsin men, parental and filial reports of status variables provide unique and direct evidence about the response biases postulated by Broom et al. There is negligible evidence of such biases. Given the relative strengths of the Australian and Wisconsin research designs, we do not believe that further concern about the presence of this type of response error is warranted.

For the past 15 years a major line of research has been based on structural equation models of the process of socioeconomic achievement (Blau and Duncan 1967; Sewell and Hauser 1975). Lately, findings in this research tradition have been questioned because they depend on survey-based, retrospective proxy reports of socioeconomic variables that may be subject to large response errors. Clearly, the failure to recognize and

compensate for errors in variables may lead to serious biases in estimated parameters of the stratification process (Bowles and Gintis 1976). However, the direction and importance of such biases depend on the size and location of random and nonrandom (correlated) response errors and on the composition of the population to which the error-ridden reports pertain (Bielby, Hauser, and Featherman 1977a; Griliches 1979). That is, the fact that survey responses are subject to error has no obvious or determinate implications for models of stratification; response variability must be incorporated into socioeconomic models.

A substantial body of evidence has been accumulated that bears on the effects of age, race, and sex on the validity of reports of socioeconomic variables (Borus and Nestel 1973; Kerckhoff, Mason, and Poss 1973; Mason et al. 1976; Mare and Mason 1980; Bielby et al. 1977a, 1977b; Corcoran 1979, 1980) as well as on the consistency of reports across multiple measurement occasions (Bielby et al. 1977b; Bielby and Hauser 1977). All of the cited studies yield estimates of validity or reliability that are moderate to high (.7–.9), and none of the studies has provided substantial evidence of nonrandom measurement error. The accumulated evidence suggests that response variability is due partly to inaccurate perception of one's background status and partly to lapses of recall; it may also be sensitive to survey instrumentation, interviewer training, coding methods, and other data-collection procedures.

THE AUSTRALIAN STUDY

With the first source of error (inaccurate perception) in mind, Broom et al. (1977, 1978) propose an alternative to the traditional methodology of stratification surveys. Such surveys have usually asked a sample of respondents to report their own social and economic characteristics and those of their parents; Broom et al. call this perspective 1 (per 1). The traditional measures of social origin are retrospective reports of parents' status characteristics when the respondent was growing up (Featherman and Hauser 1975; Featherman 1980). The alternative proposal by Broom et al. (1978, p. 419) is to ask a sample of respondents to report about themselves and about their offspring (per 2): "A father's description of his son's career should be at least as accurate and possibly more accurate than a son's report of his father's. . . . From the standpoint of knowledge and awareness of adult roles, the father seems the more reliable and accurate source if a choice must be made."

From a 1973 Australian survey, Broom et al. (1978) selected two comparable and nonoverlapping subsamples in which father and sons reported about themselves and about their sons and fathers, respectively. This made it possible to compare reports obtained under per 1 and per 2

421

between nominally comparable populations. They find that higher inter-generational correlations between status variables were obtained under per 2 than under per 1 in their Australian sample; they attribute the lower correlations observed under per 1 to complex patterns of correlated response error.

We believe that the analyses of Broom et al. warrant no conclusions about differences between per 1 and per 2 in data quality. Indeed, they acknowledge (1978, pp. 425–26) that their data are not complete enough to support a definitive analysis of differences in data quality between paternal and filial respondents.

Broom et al. observe means and variances of five status variables under each perspective, but only the 10 correlations among those five variables within each perspective (Broom et al. 1978, p. 422). This leaves out the 25 correlations among status variables across perspectives. In a preliminary analysis of the Australian data, we compared the per 1 and per 2 variances and covariances without imposing a structural model.[2] The results were consistent with the observations of Broom et al.; we found statistically significant differences both in variances and in correlations between the Australian data ascertained under per 1 and those ascertained under per 2. These primarily involve differences in the variance of son's education, the correlation between father's occupational status and son's current job, and the correlation between son's education and his first job.

Broom et al. (1978, pp. 424–25) attribute the differences between the per 1 and per 2 data to correlations between response errors within each perspective. Lacking cross-perspective correlations, they cannot estimate a full measurement model to test this idea. Instead, they argue that in the Australian sample, uncorrelated errors imply implausible reliabilities of proxy reports in each perspective. For example, in the case of father's occupation and son's occupation, seemingly plausible numeric values of the true population correlation (.550) and of the reliabilities of self-reports of observable variables (.85) imply reliabilities of .327 for father's occupation in per 1 and .984 for son's occupation in per 2. In effect, the observation of larger correlations within per 2 than within per 1 leads Broom et al. to conclude that response errors are negatively correlated within per 1 or positively correlated within per 2. With reference to other correlations in the Australian data, Broom et al. (1978, p. 425) conclude that "because equally implausible and mutually inconsistent estimates result from similar analyses of all correlations . . . measurement error in these data must be nonrandom."

In fact, we can evaluate the hypothesis that the Australian data were

[2] Our analyses were based upon the correlations in table 1 of Broom et al. (1978, p. 422) and listwise-present counts of 363 in per 1 and 344 in per 2, kindly supplied by F. L. Jones. Details are available from us on request.

generated under random response variability in a single statistical population without making any assumptions about true correlations or reliabilities. That is, without cross-perspective correlations, it is possible to generate overidentifying restrictions on the model of uncorrelated response error. These restrictions pertain to ratios of the reliabilities of per 1 and per 2 reports of each variable.

Let the measurement model be

$$x_{ij} = \omega_{ij}\xi_i + \delta_{ij} , \qquad (1)$$

where x_{ij} is the observation of the ith variable in the jth perspective, ξ_i is the true value of the ith variable, ω_{ij} is the loading of x_{ij} on ξ_i, δ_{ij} is a disturbance, and all variables have been normalized with zero means and unit variances. Under the model, $E[\xi_i\delta_{ij}] = E[\delta_{ij}\delta_{i'j'}] = 0$ unless $i = i'$ and $j = j'$. Furthermore, let the true correlation between unobservables be $E[\xi_i\xi_{i'}] = \phi_{ii'}$ and the population correlation between observables in perspective j be $E[x_{ij}x_{i'j}] = \rho_{ii'j}$. Under this model,

$$\frac{\rho_{121}}{\rho_{122}} = \frac{\omega_{11}\phi_{12}\omega_{21}}{\omega_{12}\phi_{12}\omega_{22}} = \frac{\omega_{11}\omega_{21}}{\omega_{12}\omega_{22}} = \gamma_{12} . \qquad (2)$$

Similarly, let

$$\gamma_{13} = \frac{\rho_{131}}{\rho_{132}} \qquad (3)$$

and

$$\gamma_{23} = \frac{\rho_{231}}{\rho_{232}} . \qquad (4)$$

By algebraic substitution, it is easy to show that

$$\frac{\omega_{21}}{\omega_{22}} = \left(\frac{\gamma_{12}\gamma_{23}}{\gamma_{13}}\right)^{1/2} , \qquad (5)$$

but the model yields other distinct expressions for the same ratio of loadings, for example,

$$\frac{\omega_{21}}{\omega_{22}} = \left(\frac{\gamma_{12}\gamma_{25}}{\gamma_{15}}\right)^{1/2} . \qquad (6)$$

423

In all, the model of equation (1) yields six empirically distinct estimators for each ratio of loadings of the five constructs. These are displayed in table 1. For example, the relative loadings of father's education vary from .573 to 1.038, with a ratio less than 1.0 indicating superior reliability under per 2. We have not carried out formal statistical tests of equality among the six alternative estimators of each ratio of loadings, but we believe that the differences among estimates are far too large to have occurred by chance. Most of the estimates of relative loadings suggest that fathers are more reliable respondents than sons, but we are reluctant to draw any firm conclusions from these findings.

The apparent failure of the Australian data to conform to the model of random measurement error suggests two (nonexclusive) explanations: (1) per 1 and per 2 measurements are strictly comparable with respect to population definition and coverage and with respect to methods of survey measurement and coding, but they are subject to correlated response errors; (2) the per 1 and per 2 measurements are not subject to correlated response errors, but some artifact of survey design or procedure may have invalidated the comparison between them. We prefer the latter of these explanations.

The Australian survey did not use the same items to measure socio-economic characteristics of sons in per 2 as to measure socioeconomic characteristics of respondents who appeared as sons in per 1. A highly abbreviated protocol was used to ascertain educational and occupational data for sons from paternal respondents (in per 2), whereas respondent sons (in per 1) were queried at length and in detail about their educational and occupational histories. There were less striking differences in the protocols used to obtain data about fathers in per 1 and per 2; one important difference appears to be the absence of a firm temporal referent

TABLE 1

ALTERNATIVE ESTIMATES OF RELATIVE LOADINGS IN THE AUSTRALIAN DATA

INSTRUMENTS (j, k)	TRUE VARIABLE (i)				
	1	2	3	4	5
1,2	. . .	. . .	.652	.632	.612
1,3	. . .	1.010	. . .	.785	1.011
1,4	. . .	1.277	1.024	. . .	1.108
1,5	. . .	.942	1.004	.844	. . .
2,3	.968	. . .	. . .	.993	.943
2,4 .....`.......	.766	. . .	.811	. . .	.818
2,5	1.038	. . .	1.077	1.144	. . .
3,4	.616	.812	. . .	. . .	1.086
3,5	.628	.612	. . .	.861	. . .
4,5	.573	.705	.935	. . .	. . .

SOURCE.—Broom et al. 1978, table 1.

NOTE.—Variables are 1 = father's education, 2 = father's job, 3 = son's education, 4 = son's first job, 5 = son's present job. See text for explanation of the estimates.

424

for father's occupation (relative to son's age) in per 2 (see Broom et al. 1977, appendix).

We think it unlikely that intrinsic differences between fathers and sons in the quality of survey responses account for observed differences in the measurements obtained under per 1 and per 2 in the Australian data. Rather than attempt to locate possible artifacts in the Australian instruments or procedures, we have reached this conclusion by following up the closing proposal of Broom et al. (1978, p. 426): "Only studies deliberately designed to get information from both members of father-son pairs will definitively deal with the problem of proxy versus self-reports."

THE WISCONSIN DATA

From a large Wisconsin sample of high school graduates, we have obtained measures of several socioeconomic variables that permit comparison of the quality of self- and proxy reports in respondent-parent pairs. The full Wisconsin sample consists of more than 10,000 male and female high school graduates of 1957 who have been followed for more than 20 years (Sewell and Hauser 1980). The present analysis pertains to men who lived in Wisconsin in 1964 and were employed in the civilian labor force; in 1964 roughly half the sample was in school or in the military or had moved out of the state.

Table 2 shows the source and timing of measurements. For each status variable—father's 1957 occupational status (FO), son's 1964 occupational status (OC), parents' income when the respondent was a senior (PI), and son's educational attainment (ED)—there is one self-report $(XX\text{–}SE)$ and one proxy report $(XX\text{–}PR)$. Data on occupations and income from Wisconsin tax returns are taken to be self-reports, as is a 1975 telephone interview report of the son's educational attainment. A 1964 mail survey of parents and the 1975 interview provide proxy reports of the status variables.

The questions used to ascertain occupational status differed substantially across instruments. The tax returns gave only a one-line occupational title, supplemented by name of employer. The 1964 postcard survey asked for the occupation, job title, and name of firm. In the 1975 survey, the occupation questions followed usage in the Current Population Survey: "What kind of work is that? What are the most important activities or duties? In what kind of business or industry was this? Was this mainly manufacturing, wholesale trade, retail trade, or something else?" plus the standard checkoff items for class of worker. Thus, if proxy or self-reporting were not at issue, one might expect the items in the 1975 survey to be the most valid, followed by those in the 1964 survey, and trailed by the items from tax returns. The occupation reports were scaled using

TABLE 2

DATA AND SPECIFICATION OF CORRELATED RESPONSE ERRORS: MEN FROM THE 1957 WISCONSIN HIGH SCHOOL SENIOR COHORT IN THE EXPERIENCED CIVILIAN LABOR FORCE AND LIVING IN WISCONSIN IN 1964 ($N = 1,452$)

Variables	Source	1	2	3	4	5	6	7	8
Father's occupation:									
1. Self-report (FO-SE)......	1957 tax return	1.000		a-1			b	e	b
2. Proxy report (FO-PR)....	1975 interview	.697	1.000		c-1	c-2		d-2	
Parents' income:									
3. Self-report (PI-SE)......	1957 tax return	.488	.472	1.000			b	e	b
4. Proxy report (PI-PR)....	1975 interview	.339	.325	.446	1.000	c-2		d-2	
Son's education:									
5. Self-report (ED-SE)......	1975 interview	.248	.282	.201	.065	1.000		d-1	
6. Proxy report (ED-PR)....	1964 mail survey	.242	.275	.200	.054	.811	1.000		a-2
Son's occupation:									
7. Self-report (OC-SE)......	1964 tax return	.226	.249	.182	.045	.523	.488	1.000	
8. Proxy report (OC-PR)....	1964 mail survey	.240	.266	.238	.084	.611	.605	.652	1.000
Mean...........	...	28.39	32.46	8.58	8.94	13.15	13.38	35.32	10.61
Standard deviation	...	20.64	22.04	.43	.43	1.83	1.76	23.72	23.17
Standard deviation of error...	...	10.67	12.72	.25	.34	.78	.78	16.18	10.67

NOTE.—Entries below diagonal are correlations; entries above diagonal specify types of correlated error. Standard deviations of errors are based on model B13 of table 3. Error types are: a-1 = within instrument, parent respondent, own status; a-2 = within instrument, parent respondent, son's status; b = between instrument, parent respondent, son's status; c-1 = within instrument, son respondent, parent's status; c-2 = within instrument, son respondent, parent's and own status; d-1 = between instrument, son respondent, own status; d-2 = between instrument, son respondent, parent's own status; and e = within method, between instrument, between respondent.

Duncan SEI scores for detailed census occupation, industry, and class of worker categories (Duncan 1961).[3]

Educational attainment is coded in years of schooling completed, with 17 or more years scored as 17. Some additional schooling undoubtedly occurred between 1964 and 1975. This permits true discrepancies between the parent's 1964 report of son's education and the son's report of his own educational attainment in 1975. Thus our estimates of error variances in the education measures are likely to be too large. However, much of the schooling between 1964 and 1975 probably occurred in the upper tail of the educational distribution, which is collapsed in our analysis.

Parents' income is coded in the logarithm of dollars, with the range limited to $3,000–$15,000. This scaling procedure maximized correlations between the parents' income and the sons' status.

MODELS OF PARENTAL AND FILIAL REPORTS

Correlations among the items, means, and standard deviations are shown in table 2. We estimated several measurement models using Jöreskog's general method for the analysis of covariance structures (Sörbom and Jöreskog 1981). Jöreskog's method yields maximum-likelihood estimates under multivariate normality, but this distributional assumption undoubtedly does not hold in the Wisconsin data. Table 3 displays goodness-of-fit tests for selected measurement models. The likelihood ratio test statistic (L^2) has an asymptotic χ^2 distribution with degrees of freedom (df) equal to the difference between the number of variances and covariances and the number of independent parameters in a model. It contrasts the null hypothesis that the constraints imposed on the variance-covariance matrix by the parameters of the model are satisfied in the population with the alternative that the variance-covariance matrix is unrestricted. The difference in likelihood ratio statistics between two nested models, that is, a general model and a constrained version of that model, provides a likelihood ratio test of the constraints.

The first model (row A1) specifies that response errors are strictly random and that the regression of each observed variable on its true value has the same slope. Clearly, that model must be rejected. The fit improves substantially when constraints on the relative slopes are removed (row A2). The only two slopes that do not differ significantly are those for the reports of father's occupational status (see row A3).[4]

---

[3] Our results are not substantially affected by the aggregation of occupational categories or by the inclusion of men of farm origin in the sample (Massagli and Hauser 1981, pp. 444–45).

[4] Proxy reports were chosen as the reference indicator of each true variable and thus have slopes of 1.0. In the final model (row B13 in table 3), self-reports of parents' income (1.50)

We also examined the error variances of the observables. These reflect the scatter of observations about the true values. An exhaustive examination of alternative models showed that the only equality constraints on error variances that are satisfied by these data are those between the two reports of son's education and between fathers' reports of their own and their son's occupations (row A4).[5]

We used model A4 as a baseline in our search for significant error correlations. Above the diagonal in table 2 we have displayed eight

TABLE 3

CHI-SQUARE GOODNESS-OF-FIT TESTS FOR EIGHT-VARIABLE, FOUR-FACTOR
MEASUREMENT MODEL OF BACKGROUND AND CURRENT STATUS OF WISCONSIN MEN
$(N = 1,452)$

Model	L^2	df	P
A. Uncorrelated errors:			
1. All slopes constrained equal	101.709	18	.0000
2. No slope constraints	34.151	14	.0020
3. Father's occupation slopes constrained equal	36.733	15	.0014
4. Model 3 plus error variance constraints	37.451	17	.0029
$(\theta_{11} = \theta_{42}, \theta_{31} = \theta_{32})$			
B. Correlated errors (conditional on model A4):			
1. (a-1,a-2,b,c-1,c-2,d-1,d-2,e)	16.086	3	.0011
2. (,a-2,b,c-1,c-2,d-1,d-2,e)	17.297	4	.0017
3. (a-1, ,b,c-1,c-2,d-1,d-2,e)	16.100	4	.0029
4. (a-1,a-2, ,c-1,c-2,d-1,d-2,e)	21.187	7	.0035
5. (a-1,a-2,b, ,c-2,d-1,d-2,e)	16.905	4	.0020
6. (a-1,a-2,b,c-1, ,d-1,d-2,e)	18.029	5	.0029
7. (a-1,a-2,b,c-1,c-2, ,d-2,e)	21.188	4	.0003
8. (a-1,a-2,b,c-1,c-2,d-1, ,e)	22.203	5	.0005
9. (a-1,a-2,b,c-1,c-2, , ,e)	26.346	6	.0002
10. (a-1,a-2,b,c-1,c-2,d-1,d-2,)	17.000	5	.0045
11. (d-1)	32.895	16	.0076
12. (d-2)	32.697	15	.0052
13. (d-1,d-2)	28.035	14	.0141
C. Tests for significance of type-d correlated errors:			
1. Model B7 vs. model B1	5.102	1	<.05
2. Model B8 vs. model B1	6.117	2	<.05
3. Model B9 vs. model B1	10.260	3	<.02
4. Model B11 vs. model A4	4.556	1	<.05
5. Model B12 vs. model A4	4.754	2	<.10
6. Model B13 vs. model A4	9.416	3	<.05
7. Model B11 vs. model B13	4.860	2	>.05
8. Model B12 vs. model B13	4.662	1	<.05

NOTE.—Types of correlated error are identified in table 2. These models were estimated using COFAMM, version 1, release 2, with convergence criteria for the gradient under the Fletcher and Powell iterations set at $3 \times 10E$-09 (Sörbom and Jöreskog 1976).

and son's education (1.05) have relatively steeper slopes than those of the proxy reports, while the self-report of son's occupational status (.84) has a relatively flatter slope than the proxy report.

[5] Estimated error variances under model B13 are given at the bottom of table 2. In this respect, it can be stated both that per 2 reports are not uniformly superior to per 1 reports and that self-reports are not uniformly superior to proxy reports.

distinct types of correlated error that are identified in the measurement model. These eight types consist of variations of within-instrument/within-respondent, within-instrument/between-respondent, and between-instrument/within-respondent correlated error. The within-respondent error correlations are of particular interest in a comparison of the quality of per 1 and per 2 data, given the hypothesis advanced by Broom et al. We find no substantial evidence of correlated response error. As shown in row B1 of table 3, the introduction of all eight types of correlated error does not significantly improve the fit; the contrast between models A4 and B1 yields $L^2 = 21.36$ with 14 df ($P > .05$).

Despite this overall negative finding, we continued to look for evidence of specific types of correlated response error. Models B2–10 are nested between models A4 and B1; each of them deletes one or two types of correlated response error from model B1. Only in the cases of error correlation types d-1 and d-2—which pertain to between-instrument correlations of son's report of his 1964 occupation (in the tax data) with his reports of father's occupation, parents' income, and own education (in the 1975 survey)—do any of the error correlations even marginally affect the fit of the model. However, as shown in rows C1–3 of table 3, separate or joint deletion of error correlation types d-1 and d-2 leads to statistically significant deterioration in the fit of model B1.

The marginality of these findings is suggested by inconclusive results obtained by forward selection from model A4. Error correlation types d-1 and d-2 are added separately and jointly to model A4 in rows B11–13 of table 3, and the contrasts between the latter models and model A4 are shown in rows C4–6. The earlier findings are confirmed in the case of error correlation type d-1, which improves the fit both of model A4 and of the model (B12) that incorporates error type d-2 (see row C8). At the same time, error correlation type d-2 does not significantly improve the fit of model A4, nor does it improve the fit of the model (B11) that incorporates error correlation type d-1 (see row C7).

In fact, in model B13 only the correlation between the error in the tax report of son's occupation and that in the self-report of educational attainment (.084) is statistically significant. The correlations between the former error and those in the proxy reports of father's occupation (.040) and of parents' income ($-.056$) are trivial. These error correlations do not suggest any systematic tendency for sons either to overstate or to understate the correlation between social origins and destinations, nor do they have any important effect on estimates of the true correlations among status variables. Table 4 shows correlations among the four status variables estimated under model A4 (without correlated error) and model B13 (with correlated error types d-1 and d-2). Obviously, the differences between the two sets of correlations are negligible.
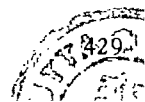
CONCLUSION

From their comparison of subsamples of paternal and filial respondents, Broom et al. (1978, p. 425) conclude, "In our case respondents report (perhaps unconsciously) their socioeconomic careers and the careers of their kin in such ways as either to maximize differences between themselves and their fathers or to minimize differences between themselves and their sons." This has been taken most seriously by at least one researcher, Hamilton (1981, p. 181), who writes that the evidence from Broom et al. "may undermine the findings of mobility research, the paradigmatic achievement of empirical sociology." On the basis of our analysis of the Wisconsin data, we do not believe that further concern about the presence of this type of correlated response error is warranted.

In no way do we suggest that "it is true what they say about Daddy." A substantial body of evidence confirms the importance of specifying response error in models of the stratification process, and the present findings are entirely consistent with that evidence. Corrected correlations are far larger than the corresponding correlations among observed variables (cf. tables 2 and 4). At the same time, we emphasize that the corrections are not necessitated by correlations between response errors of the sort postulated by Broom et al. Instead, they appear to result from simple unreliability.

TABLE 4

ESTIMATED CORRELATIONS BETWEEN LATENT VARIABLES IN THE FOUR-FACTOR MODEL OF BACKGROUND AND CURRENT STATUS ESTIMATED FOR WISCONSIN MEN

Variables	Father's Occupation	Parents' Income	Son's Education	Son's Occupation
Father's occupation	1.000	.707	.344	.350
Parents' income	.709	1.000	.246	.294
Son's education	.343	.246	1.000	.763
Son's occupation	.346	.298	.756	1.000

NOTE.—Entries above the diagonal are from model A4 in table 3; entries below the diagonal are from model B13 in table 3.

REFERENCES

Bielby, W. T., and R. M. Hauser. 1977. "Response Error in Earnings Functions for Nonblack Males." *Sociological Methods and Research* 6:241–80.
Bielby, W. T., R. M. Hauser, and D. L. Featherman. 1977a. "Response Errors of Black and Nonblack Males in Models of the Intergenerational Transmission of Socioeconomic Status." *American Journal of Sociology* 82:1242–88.
————. 1977b. "Response Errors of Non-black Males in Models of the Stratification Process." *Journal of the American Statistical Association* 72:723–35.
Blau, P. M., and O. D. Duncan. 1967. *The American Occupational Structure.* New York: Free Press.

Borus, M. E., and G. Nestel. 1973. "Response Bias in Reports of Father's Education and Socioeconomic Status." *Journal of the American Statistical Association* 68:816–20.

Bowles, S., and H. Gintis. 1976. *Schooling in Capitalist America.* New York: Basic.

Broom, L., P. Duncan-Jones, F. L. Jones, and P. McDonnell. 1977. *Investigating Social Mobility.* Departmental Monograph 1. Canberra: Department of Sociology, Research School of Social Sciences, Australian National University.

Broom, L., F. L. Jones, P. McDonnell, and P. Duncan-Jones. 1978. "Is It True What They Say about Daddy?" *American Journal of Sociology* 84:417–26.

Corcoran, M. 1979. "Measurement Error in Status Attainment Models." Pp. 269–90 in *Five Thousand American Families: Patterns of Economic Progress.* Vol. 7, *Analyses of the First Ten Years of the Panel Study of Income Dynamics,* edited by G. J. Duncan and J. N. Morgan. Ann Arbor: Institute for Social Research, University of Michigan.

———. 1980. "Sex Differences in Measurement Error in Status Attainment Models." *Sociological Methods and Research* 9:199–217.

Duncan, O. D. 1961. "A Socioeconomic Index for All Occupations." Pp. 109–38 in *Occupations and Social Status,* edited by A. J. Reiss, Jr. New York: Free Press.

Featherman, D. L. 1980. "Retrospective Longitudinal Research: Methodological Considerations." *Journal of Economics and Business* 32:152–69.

Featherman, D. L., and R. M. Hauser. 1975. "Design for a Replicate Study of Social Mobility in the United States." Pp. 219–52 in *Social Indicator Models,* edited by K. C. Land and S. Spilerman. New York: Russell Sage Foundation.

Griliches, Z. 1979. "Sibling Models and Data in Economics: Beginnings of a Survey." *Journal of Political Economy* 87, no. 5, pt. 2: S65–S98.

Hamilton, L. C. 1981. "Self Reports of Academic Performances: Response Errors Are Not Well Behaved." *Sociological Methods and Research* 10:165–85.

Kerckhoff, A. C., W. M. Mason, and S. S. Poss. 1973. "On the Accuracy of Children's Reports of Family Social Status." *Sociology of Education* 46:219–47.

Mare, R. D., and W. M. Mason. 1980. "Children's Reports of Parental Socioeconomic Status: A Multiple Group Measurement Model." *Sociological Methods and Research* 9:178–98.

Mason, W. M., R. M. Hauser, A. C. Kerckhoff, S. S. Poss, and K. Manton. 1976. "Models of Response Error in Student Reports of Parental Socioeconomic Characteristics." Pp. 443–94 in *Schooling and Achievement in American Society,* edited by W. H. Sewell, R. M. Hauser, and D. L. Featherman. New York: Academic Press.

Massagli, M. P., and R. M. Hauser. 1981. "Response Variability in Reports of Socioeconomic Data." Pp. 442–47 in *Proceedings of the Social Statistics Section, American Statistical Association.* Washington, D.C.: American Statistical Association.

Sewell, W. H., and R. M. Hauser. 1975. *Education, Occupation, and Earnings.* New York: Academic Press.

———. 1980. "The Wisconsin Longitudinal Study of Social and Psychological Factors in Aspirations and Achievements." Pp. 59–99 in *Research in Sociology of Education and Socialization.* Vol. 1. Edited by A. C. Kerckhoff. Greenwich, Conn.: JAI.

Sörbom, D., and K. G. Jöreskog. 1976. *COFAMM. Confirmatory Factor Analysis with Model Modification. A FORTRAN IV Program.* Chicago: National Educational Resources.

———. 1981. "The Use of LISREL in Sociological Model Building." Pp. 179–99 in *Factor Analysis and Measurement in Sociological Research,* edited by D. J. Jackson and E. F. Borgatta. Beverly Hills, Calif.: Sage.

Commentary and Debate

To conserve space for the publication of original contributions to scholarship, the comments in this section must be limited to brief critiques. They are expected to address specific errors or flaws in articles and reviews published in the *AJS*. Comments on articles are not to exceed 1,500 words, those on reviews 750 words. Longer or less narrowly focused critiques should be submitted as articles. Authors of articles and reviews are invited to reply to comments, keeping their replies to the length of the specific comment. The *AJS* does not publish commenters' rebuttals to authors' replies. We reserve the right to reject inappropriate or excessively minor comments.

COMMENT ON KEMPER'S "SOCIAL CONSTRUCTIONIST AND POSITIVIST APPROACHES TO THE SOCIOLOGY OF EMOTIONS"

According to Theodore Kemper (*AJS* 87 [September 1981]: 337–62), social constructionists are not interested in social structure. Only positivists are. According to him, I say that feeling rules make feeling. Discarding such simplemindedness, Kemper nonetheless tacks on the idea of feeling rules, as a finishing touch, to his own positivist model of emotion and proposes the result as the social model of emotion we have been looking for. In the first assertion, he is mistaken. In the second, he constructs a straw person in order to knock it down and to introduce, in the third, a "new dawn."

Kemper begins by distinguishing between what he calls social constructionists (those who view emotions as interpretive accomplishments of the actor) and positivists (those who view emotion as more or less jointly determined by social structure and biology). It is the positivists, according to Kemper, who lay claim to social structure. In spite of their social orientation, the constructionists apparently are not interested. Social constructionists *are*, of course, interested in social structure, but not simply in Kemper's definition of it. There are many *aspects* of social structure. One is Kemper's own provocative power-status model of interpersonal relations. Another is the pattern of relations between groups or strata

Permission to reprint a comment printed in this section may be obtained only from the author.

based, for example, on social class, sex, or race. My own entry into the domain of social structure is via the marketplace and the sale of labor that calls for emotion management—what I call emotional labor. As I say in the article Kemper criticizes, "When the manager gives the company his enthusiastic faith, when the airline stewardess gives her passengers her psyched up but quasi-genuine reassuring warmth, what is sold as an aspect of labor power is deep acting" ("Emotion Work, Feeling Rules, and Social Structure" [*AJS* 85, no. 2, 569]). A company run for profit is part of social structure. A market on which labor is sold is part of social structure. The turning of human feeling into commodities that results from the coalescence of these two has everything to do with social structure. What it does not have to do with—directly, at any rate—is Kemper's version of "social structure."

Second, according to Kemper I say that social norms "determine" or "produce" emotions; not regulate but determine and produce. Let me quote from the (fourth) paragraph, which Kemper must have skipped:

> There are two possible approaches to the social ordering of emotive experience. One is to study the social factors that induce or stimulate primary (i.e., nonreflective, though by definition conscious) emotions—emotions passively undergone. The other is to study secondary acts performed upon the ongoing nonreflective stream of primary emotive experience. The first approach focuses on how social factors affect what people feel, the second on how social factors affect what people think and do about what they feel (i.e., acts of assessment and management). Those who take the first approach might regard those who take the second as being "overly cognitive." But in fact the two approaches are compatible, and indeed the second, taken here, relies on some accumulation of knowledge garnered from the first. [P. 552]

There are many complex links between social factors and emotion, and I am asking questions about one of these links. Kemper later (p. 355) tells us, as if it were a new insight, that feeling rules are one of the ways culture affects feeling, but that there are other ways too. Right. It is irksome to have one's carefully reasoned prose straitjacketed into a discardable simplicity. But it is doubly so to then find one's own ideas rediscovered in a critique of them.

Third, Kemper ends by adding a theory of feeling rules to his positivist model of emotions in which relational outcomes varying in power and status determine the physiological states we call by emotion names. I see problems in the model to which mine has been attached and have explained them in a review essay (1981). The first problem is that: to Kemper's schema—and to mine as well—we need to add many *more* theoretical linkups to macrostructure. How might the relational complexes Kemper describes, for example, be related to the notion proposed by Michael Burawoy (1979) of the "lateralization" of conflict and, closely related to

that, the lateralization of blame and anger? That is, when incentive systems are structured in such a way as to induce workers to blame other workers rather than the constructors of the incentive system (the bosses), what alteration in the power/status pattern has led to this, and what has caused it? How then might this notion of the lateralization of conflict be tied to power and status relations, on one hand, and the interests of class and sex, on the other? Kemper's model is not yet structural enough.

The other problem lies to the other side in the need to develop a theory of the interpretation of gestures of status and power. There are not just "gestures of status." There are double-edged compliments, faint praise, ambivalent rejection—matters of interpretation. We need the shadings. We need a focus on the translations between what Kemper assumes power and status relations to be, objectively, and what any given actor actually, variably, and complexly *takes* them to be. Close up, Kemper's knowns—status and power—turn into unknowns. This does not mean that his model cannot be developed into something more convincing. It means that the way to develop it is precisely to make this translation problematic—and to develop a theory of the relation of the outer pattern to the inner interpretation of it. That is, to be a social constructionist.

<div align="right">ARLIE HOCHSCHILD</div>

University of California, Berkeley

REFERENCES

Burawoy, Michael. 1979. *The Manufacture of Consent.* Chicago: University of Chicago Press.
Hochschild, Arlie. 1979. "Emotion Work, Feeling Rules, and Social Structure." *American Journal of Sociology* 85: (November): 551–75.
———. 1981. "Power, Status and Emotion." *Contemporary Sociology* 10 (January): 73–77.
———. 1983. *The Managed Heart.* Berkeley and Los Angeles: University of California Press, in press.

THE SOCIAL CONSTRUCTION OF EMOTION: COMMENT ON KEMPER[1]

> I should almost like to say: One no more feels sorrow in one's body than one feels seeing in one's eyes. [WITTGENSTEIN]

Now that the phenomenon of human emotions has "officially" come under the sociological lens, it is perhaps not surprising that a number of polemics

have arisen. Theodore Kemper ("Social Constructionist and Positivist Approaches to the Sociology of Emotions," *AJS* 87 [September 1981]: 336–62) has, in his article, volunteered speculation concerning some of these issues and offered sharp criticisms of recent "social constructivist" writings on emotion, largely for relegating physiological (i.e., biochemical, vasomotor, visceral) correlates of emotion. After a selective presentation of studies from the field of psychology, he concludes: "the empirical data do not support the constructivist theory that radically detaches emotion from physiological underpinnings" (p. 358). Therefore sociologists should unite with biology "ultimately to build a *sociophysiology*" (p. 359). However, Kemper's critique of constructivists misconstrues, in part, this approach, and his alternative "positivist" approach which "maintains an important link with biology" (p. 337) seems of dubious (and often unexplicated) sociological utility.

To begin with, Kemper's position that "physiological cues dominate emotional attributions" (p. 340) appears as a modern-day rehashing of William James's century-old formulation: "bodily changes follow directly the perception of an exciting fact . . . our feeling of the same changes as they occur *is* the emotion" ([1890] 1950, p. 449). Insertion of physiological attendance into the "causal" sequence of persons' emotional attributions thus recalls earlier, albeit unaccounted for, phenomenological polemics. For instance, if bodily sensations constituted emotional awareness, how would persons distinguish the nebulous differences (as they would see them) between bodily sensations as emotional correlates versus bodily sensation as nonemotional homeostatic deviation or fluctuation (Irons 1895)? Alternately, if persons do not attend to lapsed bodily sensations, as is often the case, does it mean they are unemotional? One would think not. What of fearful persons who claim to be "scared senseless"? Similarly, as Wittgenstein asks: " 'Horrible fear': is it the sensations that are so horrible?" (Wittgenstein 1980, para. 148).

Related to issues of physiology is Schachter's polemical work (Schachter and Singer 1962; Nisbett and Schachter 1966), the crux of which has been to demonstrate a variable "plasticity" in persons' emotion labeling of identical bodily arousals. Kemper expounds on this work partly to criticize Shott (1979) for a literal misreading of the later Schachter study. Yet both parties seem to miss the phenomenal illogic, hence sociological relevance, of these studies from their inception. As Mandler astutely notes: "Schachter's experiments . . . [are] perhaps less interesting in an analysis of the day to day life situation of the individual. It is relatively rare that we are in a state of physiological arousal that is independent of cognitive factors" (1975, p. 91). In other words, laboratory-induced arousal sans prior cognition is of little sociological import since persons do not typically experience emotion-based arousal in cognitive vacuums. Yet it seems incredible

435

that this artifactual trivia would merit any sociological focus, since, as Schachter noted at the outset: "In nature, of course, cognitive or situational factors trigger physiological arousal, and the triggering stimulus usually imposes the label we attach to our feelings" (Nisbett and Schachter 1966, p. 228). Finally, the possibility of discrete physiological correlates to emotion, a recurrent theme in Kemper's work, seems equally sociologically uninteresting since emotional persons (a) often do not attend to physiology, and (b) when they do, it is unlikely that they discern one emotion from another by such attendance.

Moreover, Kemper does little justice to cognitively oriented constructivist writers when he states: "constructivists allow for physiological arousal—otherwise there would be nothing to label [emotional] . . ." (p. 339). In addition to misconstruing constructivist approaches, he shows here a notable disregard of cognitive writings that approximate a constructivist perspective. Many cognitively oriented writers (Lazarus 1966, 1975; Shibutani 1961; Peters 1970; Lazarus, Averill, and Opton 1970; Lazarus, Kanner, and Folkman 1980; Mandler 1975; Pritchard 1976; Averill 1980; Katz 1980) do not deny the existence of physiological correlates to emotional disturbances.[2] Rather, biological transformations are seen as a *consequence* of a prior *"appraisal"* of some environmental event. Emotion-as-appraisal refers to an environmental "meaning analysis," by persons, of a disturbing event, that is, "cognitive evaluation of environmental situations" (Mandler 1975, p. 30). These situated constructions may approximate such pressing issues as "What does it mean?" and "What shall I do?" Here nebulous physiological innervations are openly acknowledged yet relegated to heuristic considerations as they are not seen as the constitutive source of persons' emotional excitations. *Emotions*, in cognitive approaches, *are constituted by the interpretive* ("constructive") *reactions* of persons *to environmental stimuli*, often followed by, but not constituted as, physiological sensations. Coulter (1979) epitomizes the position well: "Various sensations may be bound up with syndromes of e.g., fear, anxiety, grief, and so on, but they cannot be thought of as *constituting* these emotions. . . . What distinguishes grief from remorse and disappointment from shame is not a determinate inner feeling but responses, actions, appraisals and situations in the social world" (p. 127).

Thus, in asking an angered person to give us an account of their "felt" anger we might typically expect some form of attribution (appraisal) of

---

[2] Averill (1980), who is posited by Kemper as one of those "Having largely rejected the influences of biology on emotions" (p. 343) in fact states: "I'm not suggesting that a social level of analysis is *sufficient* for an understanding of emotion, only that it is *necessary*" (p. 336). He further states: "An emotional role is built upon or incorporates elements from, one or more biological systems of behavior" (p. 337). The context of Averill's work is an attempt to introduce sociologic to traditionally mechanistic psychologies of emotion (e.g., impulses, instincts, and drives).

another's moral improprieties, for example: "he insulted me," or "she treated me unfairly" (see Garfinkel 1967, pp. 49–53; Averill 1978; Hunsaker 1981, 1982) rather than "I registered increased norepinephrine levels." Children who engage in displays of mock anger reflect culture acquisition, not attendance of physiology.

Persons can, on occasion, be angered *at* or *about* bodily sensations (e.g., pain, discomfort, etc.) but not emotions as sensations. In such instances the body qua problematic event becomes, so to speak, an environmental object: "It is nonsensical to say that an angry person is angered *by* any bodily factors except in special circumstances where these form the object of focus of his appraisal and attention" (p. 130).

Of course, physiological activation can occasionally call attention to what would otherwise be emotional unawareness (Katz 1980), as when a student "realizes" that his or her rumbling stomach reflects an upcoming exam. It can even amplify one's initial (i.e., socially constructed) sense of a disturbance. Yet these alternative possibilities necessitate the emotional persons' situated analysis of meaning. "If one feeling an emotion has any beliefs about the feeling-states present in an emotion, they will be either of two kinds: (a) that the feeling-states are due to one's beliefs and appraisals appropriate to the emotion; or (b) at least that one does not believe the feeling-states have arisen independently of one's beliefs and appraisals. . . . 'Feeling' in 'feeling an emotion' indicates that someone feels a certain way *about* something, implying the presence of a belief and some evaluation or appraisal" (Pritchard 1976, pp. 217–18).

In addition, one can be dissuaded as to the accuracy or reasonableness of one's emotional constructions (Coulter 1979). One can engage in "intrapsychic" redefinition of emotion as cognitive activity (Lazarus 1966, 1975; Lazarus et al. 1970, 1980). Yet few (if any) can directly manipulate autonomic physiological reflexes. This contrasts with Kemper's uncritical acceptance of "organismic" approaches in which innately based emotions are "automatically triggered" (p. 359).

Therefore, contrary to Kemper's assertions that, without physiological arousal ". . . there would be nothing to label" (p. 339), it might be said that there is anything and everything in meaningful social environs from which persons locate, then construct, emotional disturbances; and that any physiological attendance does not occur in vacuo. Rather, it comes about (though not always consciously) from a prior environmental signification, that is, it is a social construction.

Of course, Kemper's physiological attendance as the constitutive feature of emotion is not uncommon. Indeed such a view is centuries old. Averill (1974), whose later work Kemper criticizes, provides an insightful historical analysis of "psychophysiological symbolism" in which "extrinsic" (e.g., culturally superfluous) meanings have been embedded into the con-

cept of emotion such that there is, unwarrantedly, "the association of emotional with physiological processes on the basis of extrinsic symbolic relationships" (p. 151). He locates the association as far back as the writings of Plato. Another account locates equations of emotion with biological processes to common linguistic obfuscation:

> In one form or another, the basic categories of emotion . . . will admit of combination with the concept of feeling: thus we get "feels angry," "feels grief stricken," "feels ashamed," "feels proud," and so on. Since we can also combine the notion of "feels" with a wide range of sensation-concepts, as in "I feel tired," "she feels toothache," "he feels hungry," "she feels thirsty," "I feel pain," the stage is set for the misassimilation of emotions to sensations. . . . This has been a pervasive mistake in the history of philosophical and psychological reflection on the subject. [Coulter 1979, p. 126]

Finally, when Kemper states: ". . . there is nothing antisociological in finding that physiology plays a differentiating and crucial role in the emotions" (p. 342), no precise explication of sociological relevance is made. Similarly, he does not define what is meant by a "complete theory of emotions" which obtains from biological inclusion. More important, and as a related, practical research matter, how is sociophysiologic research accomplished? Assuming, for a moment, that physiology entered in the "causal" sequence in which persons define emotion, how is this researched from a sociological perspective? No research outline is offered.

In typically occasioned ascriptions of emotion, persons tend, as a practical matter, to construe their social environs, not their bodily processes, in the construction of emotional events. Persons define from problematic social situations, not from inner sensations, whether occasioned conflicts of meaning will lead them to feel anger or embarrassment. It is Dewey's (1895) "bear in the woods" as a constructed social object, not lapsed visceral innervation, that incites fear. In attempting to capture the everyday experiential flavor of human emotion, sociologists might glean "clues about how actors actually define situations to produce specific emotions" (Kemper, p. 337).[3] Here biology is not denied, merely relegated to wholly sociological concerns. To the extent the persons invoke physiology as an occasioned account of their situation, then biology becomes phenomenologically interesting. Yet the theoretical context in which biology is being addressed differs altogether. Thus presumptions that "sociologists might fear giving up their birthright through any accommodation with

---

[3] While Kemper criticizes constructivists for failing to specify how persons define situations to produce specific emotions (pp. 337, 359), Shott (1979), a target of this criticism, addresses exactly this issue in outlining "role-taking" emotions (embarrassment, guilt, shame). Moreover, there are a number of fine studies on embarrassment, shame, and, more recently, anger which, while being essentially compatible with constructivist approaches, address precisely such issues. See Kemper (1978, p. 32) for a partial listing of some such studies.

biology" (p. 342n) suggestively reflects reasoned theoretical differences and/or practical concerns and not necessarily academic dogma.

DEAN HUNSAKER

University of California, Santa Barbara

REFERENCES

Averill, James R. 1974. "An Analysis of Psychophysiological Symbolism and Its Influence on Theories of Emotion." *Journal for the Theory of Social Behavior* 4(2): 147–90.
———. 1978. "Anger." Pp. 1–80 in *Nebraska Symposium on Motivation*, edited by Herbert E. Howe, Jr., and Richard A. Deinstbier. Lincoln: University of Nebraska Press.
———. 1980. "A Constructivist View of Emotion." Pp. 305–39 in *Theories of Emotion*, edited by Robert E. Plutchik and Henry Kellerman. New York: Academic Press.
Coulter, Jeff. 1979. *The Social Construction of Mind*. London: Macmillan.
Dewey, John. 1895. "The Theory of Emotion II." *Psychological Review* 2:13–32.
Garfinkel, Harold. 1967. *Studies in Ethnomethodology*. Englewood Cliffs, N.J.: Prentice-Hall.
Hunsaker, Dean. 1981. "Emotions and the Situation." Paper presented at the annual meeting of the American Sociological Association, Toronto.
———. 1982. "Anger as a Social Occasion and the Social Occasions of Anger: Constructing Meaning, Purpose, and Social Control." Paper presented at the annual meeting of the American Sociological Association, San Francisco.
Irons, David. 1895. "The Physical Basis of Emotion." *Mind* 4:92–99.
James, William. (1890) 1950. *The Principles of Psychology*. Vol. 2. New York: Dover.
Katz, Judith M. 1980. "Discrepancy, Arousal and Labeling: Towards a PsychoSocial Theory of Emotion." *Sociological Inquiry* 50(2): 147–56.
Kemper, Theodore D. 1978. "Towards a Sociology of Emotions: Some Problems and Some Solutions." *American Sociologist* 13 (February): 30–41.
Lazarus, Richard S. 1966. *Psychological Stress and the Coping Process*. New York: McGraw-Hill.
———. 1975. "The Self-Regulation of Emotion." Pp. 47–67 in *Emotions—Their Parameter and Measurement*, edited by R. Levi. New York: Raven.
Lazarus, Richard S., James R. Averill, and Edward M. Opton, Jr. 1970. "Towards a Cognitive Theory of Emotion." Pp. 207–31 in *Feelings and Emotion: The Loyola Symposium*, edited by Magda B. Arnold. New York: Academic Press.
Lazarus, Richard S., Allen D. Kanner, and Susan Folkman. 1980. "Emotions: A Cognitive-Phenomenological Analysis." Pp. 187–217 in *Theories of Emotion*, edited by Robert E. Plutchik and Henry Kellerman. New York: Academic Press.
Mandler, George. 1975. *Mind and Emotion*. New York: Wiley.
Nisbett, Richard E., and Stanley Schachter. 1966. "Cognitive Manipulation of Pain." *Journal of Experimental Social Psychology* 2 (July): 227–36.
Peters, Richard S. 1970. "The Education of Emotions." Pp. 187–203 in *Feelings and Emotions: The Loyola Symposium*, edited by Magda B. Arnold. New York: Academic Press.
Pritchard, Michael S. 1976. "On Taking Emotions Seriously." *Journal for the Theory of Social Behavior* 6(2): 211–32.
Schachter, Stanley, and David Singer. 1962. "Cognitive, Social, and Physiological Determinants of Emotional State." *Psychological Review* 69: 379–99.
Shibutani, Tomatsu. 1961. *Society and Personality*. Englewood Cliffs, N.J.: Prentice-Hall.
Shott, Susan. 1979. "Emotion and Social Life: A Symbolic Interactionist Analysis." *American Journal of Sociology* 84 (May): 1317–34.
Wittgenstein, Ludwig. 1980. *Remarks on the Philosophy of Psychology*. Vol. 2. Edited by G. H. Von Wright and Heikki Nyman. Translated by C. G. Luckhard and M. A. E. Aue. Chicago: University of Chicago Press.

REPLY TO HOCHSCHILD AND HUNSAKER

Hochschild wants to have her cake and eat it too.

Indeed, she has disarmingly acknowledged "two possible approaches to the social ordering of emotive experience. One is to study the social factors that induce or stimulate primary . . . emotions. . . . The other is to study *secondary acts* performed on the ongoing nonreflective stream of primary emotive experience" (Hochschild 1979, p. 552). Furthermore, she has said that the "two approaches are compatible, and indeed, the second . . . relies on some accumulation of knowledge garnered from the first" (p. 552).

In light of this, what are we to make of Hochschild's (1979, p. 551) additional assertions? "[E]motion is governed by social rules. . . ." (p. 551). "If we consider the nature of emotion . . . we are struck by the imperial scope of social rules" (p. 551). "Why is the emotive experience of normal adults in daily life as orderly as it is? Why, *generally speaking,* do people feel gay at parties, sad at funerals, happy at weddings? This question leads us to examine conventions of feeling . . . [which] become surprising only when we imagine, by contrast, what totally unpatterned, unpredictable emotive life might actually be like at parties, funerals, weddings, and in the family or work life of normal adults" (p. 552, italics added).

Also, in case we have not appreciated the full scope of the imperialism of social rules for feeling, Hochschild informs us that her position leads us "back to the classic question of social order from a particular vantage point—that of emotion management" (p. 552).

Perhaps these are not examples of "carefully reasoned prose." They certainly give the impression that everyday life would be emotional chaos "totally unpatterned, unpredictable" (Hochschild 1979, p. 552) without social rules for feeling and that most of the emotions that people experience are the result of "management."[1] It was precisely *against* that view, among others, that my article was directed.

My point in arguing against social constructionist views of emotion was that: (1) there are limiting conditions to emotional plasticity based on physiological concomitants of emotion, an idea that many social constructionists have resisted; (2) emotions are better explained by outcomes of social relations of power and status (social structure) which directly evoke emotions than by feeling rules and emotion management; (3) when actors cognize and interpret their social situations in order to define them, they do so in terms of the power and status meanings of actors, occasions,

---

[1] According to Hochschild (1979, p. 551), emotion management is not simply the masking or suppression of supposedly inappropriate emotions or putting on fake displays of unfelt emotion. Rather, it is "deep acting" by which new authentic emotions are created.

behaviors, and objects. It is these meanings which give rise to emotions directly. When emotion management occurs, it does so (from a sociologically interesting perspective) by reformulating the relevant power and status relations so that new emotions may flow from them directly.

Hochschild thinks that I do not give her credit for being interested in social structure. Again, it may be a problem of not so carefully reasoned prose, but what, indeed, is one to think of the following: "The interactive account of emotion [Hochschild's own] points to alternate theoretical junctures—between consciousness of feeling and consciousness of feeling rules, between feeling rules and emotion work, between feeling rules and social structure" (Hochschild 1979, p. 560). Upon reflection, one may wonder just what an "alternate theoretical juncture" is. But a reasonable reading would be that "feeling rules" and "social structure" are alternative routes to understanding emotion. Actually, in what follows in her article, Hochschild does not mean that. She does take up how social structure may affect feeling rules. But my original point remains. Hochschild sees feeling rules, not social structure, as the proximal key to emotion.

As to whether or not profit-oriented firms and wage labor constitute "social structure," I believe it depends upon what content is ascribed to these sociologically vacuous terms. Weber (1947, pp. 169, 211–14, 246–48, 262), among others, says that economic markets and exchange relationships are sociologically composed mainly of power relations. Power relations between actors are important predictors of emotions. Hence if Hochschild views firms and labor through the lens of power, she is certainly working with "social structure." But it can be hoped that, before she embarks on the analysis of emotion management in such contexts, she also examines the "primary" emotions generated by the social structure of the market in the first place, that is, does a little positivist analysis at the outset.

Hochschild says that my own power-status view of social structure is "not yet structural enough" because it does not include the "patterns of relations between groups or strata based, for example, on social class, sex, or race." She has not read carefully. I wrote: "Social structure [defined in power-status terms] is a general concept. It applies at the macrolevel, as in the relations between social categories (e.g., male and female), organizations, social classes or societies; it applies also at the microlevel, where actors are individuals in interpersonal relations" (Kemper 1981a, p. 338). As for the Burawoy problem, my social structural theory of emotions does not deal with how actors are induced to blame some and not others for their difficulties. However, once such attributions are made, interpretable in power and status terms, my theory does tell what emotions are likely to ensue.

Perhaps I am guilty of claiming a "new dawn," but it does not include

feeling rules in the way Hochschild supposes. I wrote: "Cultural and social norms for feeling *do* exist. But these are epiphenomenal, pointing to the mere surface of the phenomenon, rather than to its explanatory core" (Kemper 1981*a*, p. 345, italics in original).

Finally, I welcome Hochschild's questions directed toward a further specification of my power-status approach, which Hochschild has misunderstood previously (see Kemper 1981*b*). But in asking for "a theory of the interpretation of gestures of status and power," she is wrong. What is needed is not a theory but a *codebook* containing the details (boring though they may be) of how different power and status meanings are conveyed in given cultures through language, paralanguage, proxemics, kinesics, and the like. With this we would have a compendium of cultural devices that are learned and used by actors to convey and to understand power and status meanings, whether grossly and unmistakably or in highly nuanced terms. There is less need for interpretation, both theoretically and practically, than Hochschild thinks.

Hunsaker does not like or want sociophysiology. He seems to prefer metaphysics. While I pay due respect to Wittgenstein the philosopher, Wittgenstein the sociologist has eluded my ken. Hunsaker is aware of a "recurrent theme" in my work, namely the "possibility of discrete physiological correlates to emotion." He seems ignorant, however, of exactly how I deal with this theme and what it implies. In Kemper (1978*a*, 1980, 1981*a*) I have been content to cite only the fully developed presentation. In chapters 7 and 8 of *A Social Interaction Theory of Emotions* (Kemper 1978*b*) readers can find the complete formulation of my view of sociophysiology, along with the research program, sociological utility, and so forth, that Hunsaker desires, as well as an elaboration of what a "complete theory of emotion" would contain.

What Hunsaker has read of my work he has not read carefully. He seems to want to pillory me for not being a "cognitive" theorist, so he writes of how his favored position on emotions "contrasts with Kemper's uncritical acceptance of 'organismic' approaches in which innately based emotions are 'automatically triggered.' " Hunsaker has invented a position for me. He seems to think I have omitted the "meaning analysis" that actors go through in the process of experiencing an emotion. It is instructive, then, to see what I did say: ". . . social constructionists have correctly identified the requirement that actors identify the relationship and its outcome before an emotion will be felt. But . . . they have failed to explain how actors do that" (Kemper 1981*a*, pp. 358–59). Indeed, my contribution was to specify just what it is in social relationships that actors interpret when an emotion is evoked. Furthermore, in contrast with Hunsaker's misleading partial quote (see above), I wrote: "The positivist also accepts what Hochschild (1979, p. 553) calls the 'organismic approach' to emo-

tions, in which emotions are 'automatically triggered' (p. 554) *once the social stimulus, properly defined by the actor, is in place"* (Kemper 1981*a*, p. 359, italics added).

I have paid due court to social construction where I believe it is useful. It is not useful when it fails to understand what kinds of social processes, particularly power and status relations, can be linked to emotions. It is also not useful when it seeks to ignore physiological processes in emotion: as if we could speak of emotion without acknowledging a physiological concomitant, whether or not one is conscious of it and regardless of what Wittgenstein says.

Hunsaker misconstrues Mandler's critique of Schachter's work and fails also to see the import of the previously uncritical acceptance of that work by social constructionists. If Hunsaker wants to abandon Schachter altogether, so much the better for the sociology of emotions, as I have held (Kemper 1978*b*).

Hunsaker actually knows more about the differential physiology of emotions than he first gives us reason to suspect. He tells us that a student's stomach rumblings signify (the anxiety of) an upcoming examination. Given the circumstance, why doesn't the student have bodily sensations of comfort, well-being, and satiation, as for example, those felt after an orgasm? Hunsaker should ponder that question.

Finally, Hunsaker should consider the possibility of a theory that links patterns of social relations, experienced emotions, and physiological processes in a single, overarching framework, such as I have suggested may be possible (Kemper 1978*b*). If social constructionists cannot see the sociological value or relevance of that, there is something very limited about that perspective.

THEODORE D. KEMPER

St. John's University

REFERENCES

Hochschild, Arlie R. 1979. "Emotion Work, Feeling Rules, and Social Structure." *American Journal of Sociology* 85 (November): 551–75.
Kemper, Theodore D. 1978*a*. "Toward a Sociological Theory of Emotions: Some Problems and Some Solutions." *American Sociologist* 13 (February): 30–41.
———. 1978*b*. *A Social Interaction Theory of Emotions*. New York: Wiley.
———. 1980. "Sociology, Physiology, and Emotions: Comment on Shott." *American Journal of Sociology* 85 (May): 1418–23.
———. 1981*a*. "Social Constructionist and Positivist Approaches to the Sociology of Emotions." *American Journal of Sociology* 87 (September): 336–62.
———. 1981*b*. "Power, Status and Emotions: Kemper's Reply." *Contemporary Sociology* 10 (November): 719–21.
Weber, Max. 1947. *Theory of Social and Economic Organization*. Translated by A. M. Henderson and Talcott Parsons. New York: Free Press.

A METHODOLOGICAL COMMENT ON "ECONOMIC
SEGMENTATION, WORKER POWER, AND INCOME
INEQUALITY"

The recent paper by Kalleberg, Wallace, and Althauser ("Economic Seg-
mentation, Worker Power, and Income Inequality," *AJS* 87 [November
1981]: 651–83) is likely to have a lasting impact on the study of social
stratification in sociology. It represents the best extant attempt to combine
traditional status-attainment questions with current insights from labor
market, Marxist, and organizational approaches to stratification. Because
this paper is likely to be important, some comment is necessary on a
serious methodological flaw that weakens the authors' conclusions and
threatens the future usefulness of "structural-level" stratification research.

Because Kalleberg et al. (1981) combine individual- and structural-
level data within multiple regression models, serious problems arise in
interpreting those models. The authors use two very different sources of
information within one structural equation model. They combine aggre-
gate industry-level data with self-reported income and personal and job
characteristics. For stratification researchers like me, who are interested
in incorporating structural properties into our explanation of individual
phenomena, the appropriateness and interpretation of models such as
Kalleberg, Wallace, and Althauser provide must be understood.

Primarily the authors' data come from survey instruments familiar to
sociologists in which a sample of individuals are asked about their in-
dustry, income, occupation, and certain work characteristics. The authors
supplement this information with aggregate data (familiar to economists)
on industry averages for characteristics that are theoretically interesting.
In addition, job skill characteristics from the Dictionary of Occupational
Titles are appended to self-reported information on occupational char-
acteristics. These are widely different levels of data collection. The au-
thors admit (p. 655) that firm rather than industry data are more appropriate
theoretically, but because only industry-level data are available, they opt
for industrial averages for their measures of economic segmentation. The
authors are in a theoretical bind. They want to prove that their structural
model is superior to the individual-level analysis of traditional status-
attainment models but do not have data on the actual institutional struc-
tures which affect the occupational experience of their sample of workers.
They are forced to substitute industry averages for the actual firm- or
establishment-level data that would directly reflect the individual's struc-
tural position.

The regression analyses are carried out on an individual rather than
industry level of analysis. Individuals are assigned industry scores, and
the familiar status-attainment models are run with industry segmentation

and worker power variables included. This means that, for any one individual, the industry segmentation values assigned will be industry averages rather than the actual value for the firm (or, better still, department: see Talbert and Bose [1977]) in which the person works. The problem within a regression model becomes one of comparability between individually accurate and industry average variables in terms of the meanings of their coefficients.

Because the aggregate industry variables are industry averages, we can expect that for any one individual the error between the industry score and the true score will be much larger than the error between self-reported scores and true scores. Thus, we can expect that the economic segmentation coefficient will be unstable because of unknown and nonsystematic error components, and that these variables may tend toward nonsignificance when in fact they should be significant. In fact, an inspection of table 2 in the Kalleberg et al. article reveals that the self-reported segmentation variable (establishment size) coefficients are consistently larger and more likely to be significant than the other segmentation variables.

In trying to interpret the relative importance of different variable effects within an equation, the researcher cannot, given the available data, distinguish among differences in effect. The authors' observation that worker power contributes a larger increment to explained variance (p. 699) and the consistently more significant and generally larger coefficients of the worker power versus economic segmentation variables must be understood within this context of noncomparability.

I shall propose two methods for dealing with this issue of interpreting coefficients.

The first allows the researcher to remain at an individual level of analysis but sacrifices much information. Given a model with noncomparable independent or predictor variables—in this case, caused by vast inconsistencies in the plausible error assumptions—no comparison of coefficients should be attempted, and the investigator should rely on the significance (or nonsignificance) of different variables and their direction of effect in testing theoretical propositions. This is then similar to a log-linear analysis in the amount of information the researcher can infer from the data. In the case of the Kalleberg et al. models, this will lead to conservative interpretations of the data because the economic segmentation variables will tend toward insignificance because of their "measurement distance" from the individual worker. Thus the very existence of a significant relationship in the predicted direction, given the limitations of the data, constitutes strong confirmation and suggests that, given appropriate data, the relationship might be even stronger. Conversely, negative findings should be viewed with suspicion because of the difficulties in interpreting the actual effect and the measurement error effect. The

likelihood of Type I error is minimized while the likelihood of Type II error is increased. The general implication of this observation is that existing research on labor market segmentation that has failed to find empirical confirmation for the theoretical propositions may be an artifact of mixed levels of analysis (see Baron and Bielby [1980] for some of these disconfirmations). In the absence of firm-level data, a thorough test of structural effects on the process of individual stratification proposed by labor market theory is not possible. Approximations of firm-level data with industry averages will produce unreliable results when combined with individually specific data.

An alternate approach to the measurement problem is to abandon individual-level statistical analysis. Using this approach, the authors could have substituted industry averages for all significant variables and then run their regressions on an industrial (rather than individual) level of analysis.[1] Some of these industry-level data are already available through various Department of Commerce sources. Other data would have to be estimated by aggregating existing individual-level data from, for example, the Census of Population, Current Population Survey, or specially tailored surveys such as the Quality of Employment Survey and the Survey of Working Conditions.

DONALD TOMASKOVIC-DEVEY

Boston University

REFERENCES

Baron, J., and W. Bielby. 1980. "Bringing the Firms Back In: Stratification, Segmentation, and the Organization of Work." *American Sociological Review* 45:737–65.
Birch, D. 1979. *The Job Generation Process.* Cambridge, Mass.: M.I.T. Neighborhood and Regional Change Project.
Talbert, J., and C. E. Bose. 1977. "Wage-Attainment Processes: The Retail Clerk Case." *American Journal of Sociology* 83:403–24.

[1] An even more appropriate analysis theoretically would be on the establishment level: however, this type of data is scarce in the social sciences (but see Birch 1979; Talbert and Bose 1977; Baron and Bielby 1980).

STRUCTURAL EFFECTS ON INDIVIDUAL INEQUALITY: REPLY TO TOMASKOVIC-DEVEY

Ever since social scientists have sought to explain individual-level phenomena by using contextual measures, there have been those who have misperceived the intent or implications of cross-level models (see Cronbach 1976; Althauser, Smith, and Kalleberg 1982). In the latest recurrence of this confusion, Tomaskovic-Devey raises several issues concerning the relationship between structural properties and individual inequality. His

comments are directed at our paper (Kalleberg, Wallace, and Althauser, "Economic Segmentation, Worker Power, and Income Inequality," *AJS* 87 [November 1981]: 651–83) in which we estimated individual-level models of income inequality that incorporate concepts from market, class, and organizational explanations of stratification. Unfortunately, he confounds questions of theory or substance and methodology. These need to be clarified before any progress can be made in understanding structural sources of inequality.

Tomaskovic-Devey is mainly concerned with problems of interpreting regression coefficients in equations containing both individual and structural variables. He argues that "industry average" and "individually accurate" measures are "noncomparable" and differ in their "meanings." This is purportedly because there are greater differences between true and observed scores for industry- than for individual-level variables. As a result, the coefficients of the industry variables will be unstable because of "unknown and nonsystematic [*sic;* "systematic"?] error components." These "vast differences in the plausible error assumptions" also produce results which are "artifacts of mixed levels of analysis." He argues that our use of regression coefficients to assess relative effects of variables at multiple levels of analysis exhibits a "serious methodological flaw" which threatens not only the validity of our results but also the course of future "structural-level" stratification research.

If valid, these would indeed be serious criticisms. However, a careful consideration of the theoretical and substantive issues involved in our analysis sheds light on Tomaskovic-Devey's methodological confusion. His major theoretical error is that he assumes the "closer" a measure of structure is to the individual worker, the more appropriate it is as an explanation of income inequality. Thus, he argues that the firm is more appropriate than industry, though departments within firms, having less "measurement distance" from the worker, would be even better. We agree that firm- and department-levels of analysis are important for understanding structural sources of income inequality, and we regret that we could not measure more precisely some important firm-level concepts.[1] However, firms and departments are important because they represent structural locations within which income determination processes theoretically operate, not because they are "closer" to the worker than is industry in terms of "measurement distance." Further, firms and departments do not exhaust the range of structural sources of income inequality:

[1] We have collected such data, along with colleagues at Indiana University, as part of the 1982 Indianapolis/Tokyo Work Commitment Study. These data consist of questionnaires filled out by over 7,000 workers in almost 100 plants in seven manufacturing industries in the two countries. In addition, we collected detailed data on industry, firm, and departmental concepts through interviews with key informants in each plant. This data set will allow us to examine organizational and other structural determinants of work outcomes with much more precision than was possible in the past.

we argue that such structures operate at *many* different levels. Economic inequality is attributable not only to department and firm characteristics but to industrial, class, and occupational forces as well. These layers of structure need to be conceptualized in relation to one another and their structural effects on income analyzed accordingly.

Within our original framework, industry is conceptually distinct from "firm," and each has theoretically distinct effects on income inequality. High-wage firms may exist in low-wage industries and vice versa. While we would have preferred to measure *some* variables also at the firm level (profits, capital intensity, sales to the government), this does not detract from their utility as measures of industrial contexts. Moreover, other industrial variables (e.g., concentration) are defined *only* at the industry level because they refer to relations among firms and, therefore, have no analogue at the firm level. Hence, industry-level concepts are important in their own right and should be included in income determination models along with firm-level measures.

These theoretical considerations are needed to make sense out of the methodological issues raised. Tomaskovic-Devey's arguments regarding measurement error, for example, make sense only if we assume that measures of "distant" levels of structure are taken as proxies for "closer" levels, a point we rejected above. Further, we did not use or interpret industry measures as proxies for firm characteristics. Instead, we argued that they represent measures of industrial contexts themselves, which are conceptually distinct from firm properties.

These theoretical considerations aside, we still have problems with Tomaskovic-Devey's arguments regarding measurement bias in our models. While our industrial-level concepts are undoubtedly measured with some error, this can be said for *all* of our structural measures (or for almost all variables in the social sciences). Individual perceptions of establishment size, averages of occupational skill levels, and individual reports of class position are all contaminated by measurement error, as are even individuals' descriptions of their demographic characteristics. Tomaskovic-Devey's assertion that our measures of industry characteristics have greater (systematic) measurement error than our other variables simply has no factual basis. The industrial variables, being averages, may not describe precisely the detailed industrial contexts experienced by particular workers in our sample. But by the same token, the occupational skill measures used by us and others do not convey exactly the skill content of specific jobs. What, then, are the implications of measurement error for our analyses? Such errors will, of course, bias our results, as would be the case in any regression model. Nevertheless, it follows from this neither that the individual-level variables are measured without error ("individually accurate" measures) nor that industrial and individual variables are "noncomparable."

His arguments concerning the "noncomparability" of industry averages and individual-level variables reflect considerable confusion about the analysis of multilevel data. We *can* "compare" coefficients of industrial measures to the others in the model, provided we keep in mind the implications of such comparisons. That is, regression coefficients are useful measures of effect within a properly specified model; and "models which include contextual variables neither require special advocacy nor bear any special burden of proof" (Farkas 1974, p. 357). We can conclude from our analysis, for example, that some economic segmentation variables, such as concentration, are more strongly related to income inequality than others. Any comparison of industry and other coefficients, however, must take into account the differences between them. Since the industry averages vary only among individuals who are members of different industry groups, they will generally have smaller effects than the individual-level variables, which vary both between and within industries. We reported in the paper that the industry categories account for a maximum of 11.1% of the total variation in (log) income for males and 14.1% for females. Thus, any particular scaling of these categories cannot exceed this between-industry variation in income. This tends to make the estimates of industry effects nonsignificant.[2] In addition, since industrial characteristics are the most "distant" structural concepts in our models, their effects may be mediated by firm and other intervening variables. Thus, the fact that our industry effects are relatively small does not imply they are unimportant or even "conservative." Instead, in addition to their direct effects, industry variables affect income indirectly by determining, in part, the occupational skill mix, union structure, and other labor market structures which, in turn, affect income. That many of our estimates of industry effects are significant despite these theoretical considerations only serves to bolster our argument regarding the importance of these dimensions of economic segmentation.

Tomaskovic-Devey also suggests that we aggregate all our measures to the industry level to make them "comparable," a suggestion which we explicitly reject. This technique is useful only for answering questions for which industry can be justified as the appropriate unit of analysis.[3] However, such an approach is seriously flawed for addressing the issues which motivated our analyses in the article in question. Theoretically, our con-

[2] We could, of course, increase the amount of between-industry variance in the sample if we used industry-level data assigned to a more detailed classification than two-digit SIC codes. Despite our advocacy of more detailed industry data (see Wallace and Kalleberg 1981), such data are not generally available for the entire economy.

[3] In an earlier paper (Wallace and Kalleberg 1981), two of us used an industry-level analysis to investigate interrelations between industrial characteristics and labor market structures. However, in that paper we were explicitly concerned with institutional processes which could theoretically be justified as operating at the industry level, and not directly with inequalities among individuals.

449

cern is the explanation of income inequality among *individuals,* and the arguments are specified at the individual level of analysis. To aggregate all variables to the industry level makes a heroic (but incorrect) theoretical assumption: that this is the level at which class, organizational, and market processes generate income inequality among individuals.

Aggregating our variables to the industry level of analysis would not only contradict Tomaskovic-Devey's own arguments regarding the theoretical primacy of firms and their structural effects, it would also go against much literature which warns against the perils of making theoretical inferences at incorrect levels of analysis. In addition to inviting problems of "aggregation bias" (see Hannan, Freeman, and Meyer 1976), such aggregation would also preclude our assessment of the economic effects of class, occupational, organizational, and individual variables *within* industries. That is, since many of our variables (e.g., sex, education, skills) exhibit greater within- than between-industry variation, the inclusion of these variables in an industry-level model of income inequality would be virtually meaningless. Hence, estimating models at the industrial level, as Tomaskovic-Devey suggests doing, implies that one is concerned with explaining that portion of the 11.1%–14.1% of the total variation that lies between industries and not the other 86%–89% that lies within.

We acknowledge that because of data limitations our empirical analysis of the theoretical questions we raise is not ideal. Our analyses were based on data collected through household surveys of the general working population, which are not as well suited to examining some questions as are data collected at the firm level. Rich firm-level data are not generally available, but researchers concerned with structuralist arguments are moving in the right direction (see, e.g., Baron and Bielby 1980; n. 1). If we had ideal data at multiple levels of analysis (e.g., industry, firm, department, work group, etc.), we would still use the same analytic framework we employed in our paper. Indeed, under such conditions the utility of our approach would be even more apparent. Although it is important to learn how much firm-specific processes affect individual inequality, for example, it is still necessary to ask whether these processes mediate industry-level ones. We would also want to assess effects on income inequality produced by occupational and class concepts. Such analyses require the use of multilevel models. The analytic tools for examining these questions are straightforward and, by now, well-known. What is needed, however, is a broader understanding of how unresolved theoretical issues are distinct from and often prior to methodological questions.

ARNE L. KALLEBERG, MICHAEL WALLACE,
and ROBERT P. ALTHAUSER

Indiana University

REFERENCES

Althauser, Robert P., Kent W. Smith, and Arne L. Kalleberg. 1982. "Total, Within-, and Between-Group Coefficients: Their Multivariate Relationship." *Political Methodology* 8:1–38.

Baron, James N., and William T. Bielby. 1980. "Bringing the Firms Back In: Stratification, Segmentation, and the Organization of Work." *American Sociological Review* 45:737–65.

Cronbach, Lee J. 1976. "Research on Classrooms and Schools: Formulation of Questions, Design and Analysis." Occasional paper of the Stanford Evaluation Consortium.

Farkas, George. 1974. "Specification, Residuals and Contextual Effects." *Sociological Methods and Research* 2:333–63.

Hannan, Michael T., John H. Freeman, and John W. Meyer. 1976. "Specification of Models for Organizational Effectiveness: A Comment on Bidwell and Kasarda." *American Sociological Review* 41:136–43.

Wallace, Michael, and Arne L. Kalleberg. 1981. "Economic Organization of Firms and Labor Force Consequences: Toward a Specification of Dual Economy Theory." Pp. 77–117 in *Sociological Perspectives on Labor Markets*, edited by Ivar Berg. New York: Academic Press.

Review Essay: Is Marxism Really Functionalist, Class Reductionist, and Teleological?[1]

A Contemporary Critique of Historical Materialism. Vol. 1. *Power, Property and the State.* By Anthony Giddens. Berkeley and Los Angeles: University of California Press, 1981. Pp. xii + 294. $24.50 (cloth); $10.95 (paper).

Erik Olin Wright
University of Wisconsin—Madison

In the early days of the contemporary revival of Marxist thought, virtually the only critiques of Marxism came from the Right. Of course, not all leftist or critical intellectuals identified with Marxism, but few were openly hostile to the new directions of Marxist thought, and most saw in the new outpouring of Marxist writing great promise for theoretical clarification and advance. Marxists, for their part, were self-confident, certain of the basic soundness of their theoretical posture, and energetic in their attempts at reconstructing and developing aspects of Marxist theory that were seen as inadequate.

The situation is quite different today. One finds innumerable references to the "crisis of Marxism." In part this is, of course, a reflection of the political impasse experienced by the Left in many countries after the optimism of the late 1960s. But the sense of crisis has also been nurtured by developments in theory. On the one hand, many Marxists are more cautious in their theoretical claims, acknowledging problems and dilemmas within Marxist theory. On the other hand, there has emerged a growing body of leftist intellectual work which is highly critical of Marxism and often explicitly anti-Marxist.[2]

Two characteristics of these new critiques of Marxism are particularly important. First, they are critiques on the Left, not from the antisocialist Right. The criticisms are not from apostate Marxists who have become defenders of capitalism; they are from anticapitalist intellectuals with commitments to progressive social change. In some cases, in fact, these theorists' vision of the alternative to capitalism is not radically different from the image of socialism and communism contained in Marxist theory; what is different is the view of the theory of society needed to help create such a society. Second, the critiques are not simply critiques of the insufficiencies or gaps in Marxist theory; they are critiques of Marxism. In

one way or another all of these theorists argue that Marxist theory is a hindrance, that its theoretical assumptions necessarily create blind spots, that its foundations are fundamentally flawed and thus it cannot be reconstructed—it must be abandoned.

Anthony Giddens's recent book, *A Contemporary Critique of Historical Materialism*, holds a unique place among these new critiques of Marxism. Unlike most of the other recent critiques, Giddens's does not reflect a disillusionment with Marxism, since Giddens never embraced Marxist theory; instead, it grows out of his long-standing theoretical project of developing a critical sociology. While Giddens does see his analysis bearing on an emancipatory political project—"furthering . . . forms of human society in which the mass of human beings can attain freedom and modes of self-realization in excess of any they may have enjoyed before" (p. 24)—he is not writing primarily for an audience engaged in progressive politics; he writes as an academic sociologist primarily for sociologists. The book is thus less passionate and political than the others; but it is also, in many ways, the most serious and systematic theoretical engagement with the basic ideas of Marxism of any of these recent works. Although, as I shall attempt to show, I think many of Giddens's specific arguments against historical materialism are unsatisfactory, the book is an important contribution to debates over Marxism and deserves a serious reading by both Marxists and non-Marxists.

The core of the book deals with three general problems in Marxist theory: (1) the strong tendency in Marxist writing to adopt functionalist arguments in analyzing the social totality, (2) the tendency toward class or economic reductionism in Marxist typologies of historical forms of society, and (3) the use of evolutionary and teleological reasoning in the explanation of social change. In place of each of these errors. Giddens proposes the basic elements of an alternative: functionalism should be replaced by the analysis of society as a contingently reproduced social system, class reductionism should be replaced by a typology of social forms based on a multidimensional concept of "space-time distanciation," and evolutionism should be replaced by what Giddens calls the theory of "episodic transition."

In this short essay I cannot review the details of each of Giddens's criticisms and the alternatives he proposes. What I will do is very briefly outline the central point of each argument and suggest some of the weaknesses in Giddens's position.

Functionalism

Giddens argues correctly that a good deal of Marxist work adopts functionalist forms of argument. To take just one familiar example, the form of the state is often explained by Marxists in terms of the functional requirements of reproducing the system of class domination and exploitation. Such functionalism Giddens criticizes on a variety of familiar grounds: functional explanations rest on a false dichotomy between statics

and dynamics; they tend to turn human actors into mere bearers of relations; and, most important for Giddens, they falsely impute "needs" to social systems. In Giddens's view, the only legitimate way to use functional arguments in social science is a strictly counterfactual one. Counterfactual uses of functional arguments, however, can never be construed as explanations of anything; they simply point the way to what needs explaining.

While I think the basic thrust of Giddens's critique of functionalism, in both its Marxist and non-Marxist incarnations, is correct, his discussion is misleading in certain respects. First, he basically ignores the considerable debate over functional arguments within the Marxist tradition. He argues as if functionalism were a largely unrecognized problem in Marxist theory, whereas in fact the critique of functionalist tendencies in Marxism has been at the heart of recent debates.[3]

Second, Giddens treats the problem of functional explanation as if it were a matter to be settled on methodological grounds, rather than a substantive theoretical problem. That is, functionality, in Giddens's view, is always illegitimate as an explanation, and therefore in any specific case no substantive arguments against a proposed functional explanation are needed. One should be, I believe, highly suspicious of functional explanations, since in general they involve a rather facile shift from a functional description (the state is functional for the ruling class) to a functional explanation (the form of the state is actually explained by its functional effects). But this skepticism is the result of the absence from such arguments of any convincing feedback mechanisms which could plausibly establish the functional relation; it should not be the result of an a priori rejection of the possibility of such mechanisms. Indeed, as Elster (1979, p. 31) has argued in an analysis of profit-maximizing practices engaged in by capitalist firms, there are special cases in which genuine functional explanations can be established. Marxists have certainly been guilty of using functional arguments without a satisfactory account of mechanisms, and thus such arguments are at best incomplete and in most cases probably incorrect. But this should be argued substantively in each case rather than settled categorically by a methodological injunction against functional explanations.[4]

---

[3] See, in particular, the debate over G. A. Cohen's (1978) defense of functionalism (e.g., Elster 1980, 1982; J. Cohen 1982) and the many critiques of functionalism in the work of Poulantzas, Therborn, Althusser, and other "structuralist" Marxists (e.g., Esping-Anderson, Friedland, and Wright 1976; Jessop 1977).

[4] Giddens bases his categorical rejection of functional explanations on the claim that they always imply the attribution of a "need" to a system. While functional arguments may use the idiom of "needs" following an organic metaphor, there is really nothing in the structure of such explanations that relies on the attribution of needs. As Elster (1979, 1980, 1982) has argued in his debate with Cohen (1978, 1980, 1982), the essential property of a functional explanation is that reproductive or beneficial effects constitute explanations of their causes through a feedback mechanism that does not operate strictly through the conscious intentions of actors.

Typologies of Social Forms

In one way or another, all Marxists construct typologies of social forms based on the social organization of production and the class structure rooted in that social organization. Much of Giddens's book is devoted to challenging this way of distinguishing among societies. The accusation of class reductionism is, of course, a familiar criticism of Marxism. What is unusual about Giddens's argument is that, while he rejects general class-based typologies of societies, he defends the importance of class analysis, and, in particular, he accepts the essential Marxist characterization of capitalism as structured above all by its distinctive class relations.

The heart of his argument is that it is only in capitalist society that class constitutes the underlying structural principle of the whole society. While various kinds of noncapitalist society had classes, only in capitalism does class permeate and structure all aspects of social life. Giddens thus draws the distinction between "class-divided societies" (societies "within which there are classes, but where class analysis does not serve as the basis for identifying the basic structural principle of that society" [p. 108]) and "class society."

Giddens's defense of this claim is complex and rests on his definitions and analysis of the concepts of power, resources, and domination. The gist of the argument is that only in capitalism are the relations of domination over allocative resources (basically, in less technical terms, economic resources) the central relations that sustain power relations in general, whereas in noncapitalist societies the relations of domination over authoritative resources (basically, social-political resources) constitute the basis of power. If we identify the concept of class strictly with the relations of domination over allocative resources, it follows that a general class typology of social forms is unsatisfactory.

There are two problems with Giddens's critique. First, even in his own analysis, when he attempts to identify what it is in the structures of feudalism and capitalism that explains why in capitalism allocative resources are of primary importance in domination relations whereas in feudalism authoritative resources are of primary importance, he emphasizes the importance of the differences between property relations in the two kinds of societies. It is the institutional form of property relations in capitalism (the institutional separation of the state from production) that explains the centrality of allocative relations of domination in capitalism, whereas the fusion of authoritative and allocative relations of domination within feudal property relations explains the centrality of authoritative domination in feudalism. This comes very close to traditional Marxist explanations for the use of extra-economic coercion in feudal exploitation compared with the purely economic mechanisms of exploitation in capitalism. Thus there lurks behind Giddens's own analysis a kind of property-relations typology of social forms which is rather like Marxist accounts.

A second difficulty presented by Giddens's critique of Marxist class reductionism lies in his definition of class itself. Giddens insists that the concept of class be narrowly linked to the relations of domination "created by private ownership of property" (p. 107), with "ownership" taken to mean the direct control over the use and disposition of property and "private" designating legally guaranteed individual or family rights of disposition of that property. This means that, wherever the appropriation of surplus rests on directly coercive means, such appropriation is treated by Giddens as a result of the control over authoritative resources (e.g., military personnel) and not ownership of private property and thus not class relations.

Many Marxists, myself included, define classes in terms of the mechanisms through which surplus products or surplus labor is appropriated, not by private ownership of the means of production as such. Such appropriation always involves specific combinations of economic and political mechanisms: in feudalism exploitation involves the direct use of extra-economic coercion; in capitalism the political face of class relations is restricted to the guarantee of contracts and the supervision of the labor process. In both types of society, however, it is the mechanisms of surplus appropriation which specify the distinctive character of class relations.

The basic error in Giddens's critique, therefore, is his implicit equation of class reductionism with economic reductionism. The Marxist typology of social forms is indeed a class-based typology, but this is not reducible to a simple one-dimensional economic typology. When class relations are defined in terms of mechanisms of exploitation—which I think is the central meaning of class in the Marxist tradition—Giddens's distinction between "class society" and "class-divided society" becomes a distinction between two kinds of class societies: one in which exploitation occurs through a primarily economic mechanism and one in which exploitation involves direct coercion.

Evolutionism

One of Giddens's central objectives in his critique of Marxism is to reject all forms of evolutionary thinking in social theory. He does so for two principal reasons, one methodological, the other empirical. Methodologically, Giddens argues, evolutionary theories are based on some notion of adaptation. But societies are not organisms and thus "the idea of adaptation falls in the same category as the functional 'needs' to which I have already objected. Societies have no need to 'adapt' to (master, conquer) their material environments" (p. 21). Such arguments must be rejected as teleological.

An alternative, of course, is to reconstruct an evolutionary trajectory for social development on the basis of arguments about individual adaptation. Such a reconstruction, Giddens insists, must fail empirically, since there are no transhistorical tendencies for individual human beings

to try to improve their material conditions of existence (or any other transhistorical principles of human adaptation), and thus there can be no possibility of a general theory of social change, of historical development.

In place of such a general theory of history, Giddens proposes much more contextually limited theories of specific historical transition. Each type of society generates historically specific tendencies of development with historically specific directionalities. These can be characterized as "episodic transitions," to use Giddens's expression, but they fit into no overarching theory of the history of episodes.

This part of Giddens's argument poses, I feel, the most sustained challenge to traditional Marxism. While some Marxists (e.g., Balibar 1970) have also rejected the possibility of a general theory of transition, and thus, in effect, a Marxist theory of history, most Marxists continue to support at least the possibility of a real theory of history, a theory of the overall trajectory of historical development. This is what Giddens at least claims to reject in his critique of evolutionism.

Giddens's condemnation of evolutionism can be faulted on three grounds: first, he is wrong to insist that evolutionary theory is necessarily teleological; second, nonteleological evolutionary theory does not have the methodological flaws he attributes to it; third, his own theory of time-space distanciation is really a variety of evolutionary theory.[5] The challenge Giddens poses, therefore, is between substantively contending evolutionary theories—the class-based theory of history in Marxism and the more multidimensional resource-domination theory in Giddens—rather than between antievolutionism and evolutionism.

Although theories of social evolution, including those in the Marxist tradition, are often teleological, evolutionary theory need not posit a final state or end of development toward which history inexorably moves. Biological evolutionism does not posit the inevitability of the specific trajectory of biological change that led to *Homo sapiens;* it merely posits a developmental logic which will produce that specific trajectory under a given contingent sequence of ecological pressures and events. For a social theory to be "evolutionary" it is sufficient that social forms can be ordered in such a way that (*a*) there is a greater probability of staying at a given level than of regressing to a lower level (i.e., movements are "sticky downward" in the hierarchy of social forms), and (*b*) there is some positive probability of moving forward to higher levels of the typology. Such a theory may contain multiple trajectories; there may be "branches" and "dead ends" and even regressions. What makes it evolutionary is that

---

[5] "Time-space distanciation" refers to the degree of control of a given type of social resource—allocative or authoritative—over time and space. This is easiest to understand in terms of allocative (economic) resources: in hunting-and-gathering societies, control over allocative resources is very limited in both time and space; in industrial capitalism, such control is extended over many years through long-term planning by large corporations and the state, and over the entire globe through trade and the organizational development of the multi-national corporation.

social forms can be systematically ordered and there is a determinate structure to the pattern of possible trajectories.[6]

Now, it may or may not be possible to formulate a satisfactory theory of history in these terms, but neither success nor failure would occur for methodological reasons. Such a theory does not reify society, does not posit for society as such quasi-organismic "needs" or "drives" for adaptation. Both the Marxist theory of history, when rid of its Hegelian teleological tendencies, and Giddens's theory of time-space distanciation can be considered types of evolutionary theory in this sense.

What is at stake in the debate, therefore, are alternative accounts of the underlying logic of the trajectory or possible trajectories of historical development. Giddens characterizes such trajectories in terms of the autonomous dynamics of control over both allocative and authoritative resources. There is thus a dual logic to evolutionary development: history is animated by the autonomous impulses for the expansion of time-space distanciation with respect to allocative and to authoritative resources. Since there is no general priority of one or the other of these, their interconnection is best characterized as historically specific and contingent. Marxists, however, argue that this "dual logic" is really unified within the concepts of production, exploitation, and class. The two dimensions of domination are not autonomous but systematically linked through the social organization of production and the class relations that organization sustains. This need not suggest a unilinear trajectory of historical development, but it does suggest a more determinate theory of history than Giddens believes possible.

Giddens's book is a rich and challenging study. Even though I think his tendency to formulate substantive debates in terms of contending methodological or epistemological stances at times obscures the real issues, nevertheless, his arguments are generally systematic and rigorous and establish a productive terrain for subsequent discussion.

REFERENCES

Albert, Michael, and Robin Hahnel. 1978. *UnOrthodox Marxism: An Essay on Capitalism, Socialism and Revolution.* Boston: South End Press.
———. 1981. *Marxism and Socialist Theory.* Boston: South End Press.
Aronowitz, Stanley. 1981. *The Crisis of Historical Materialism.* New York: Praeger.
Balbus, Isaac. 1982. *Marxism and Domination: A Neo-Hegelian, Feminist, Psychoanalytic Theory of Sexual, Political, and Technological Liberation.* Princeton, N.J.: Princeton University Press.
Balibar, Etienne. 1970. "Elements for a Theory of Transition." Pp. 273–308 in *Reading Capital,* by Louis Althusser and Etienne Balibar. London: New Left Books, 1970.
Cohen, G. A. 1978. *Karl Marx's Theory of History: A Defense.* Princeton, N.J.: Princeton University Press.
———. 1980. "Functional Explanation: A Reply to Elster." *Political Studies* 28:129–35.

[6] For a more extended discussion of this conceptualization of evolutionary theory, see Wright (1983).

———. 1982. "Reply to Elster on Marxism, Functionalism and Game Theory." *Theory and Society* 11:483–96.

Cohen, Joshua. 1982. "Review of *Karl Marx's Theory of History: A Defense.*" *Journal of Philosophy* 79:253–73.

Cutler, Anthony, Barry Hindess, Paul Q. Hirst, and Athar Hussain. 1979. *Marx's Capital and Capitalism Today.* Vol. 1. London: Routledge & Kegan Paul.

———. 1980. *Marx's Capital and Capitalism Today.* Vol. 2. London: Routledge & Kegan Paul.

Elster, Jon. 1979. *Ulysses and the Sirens.* Cambridge: Cambridge University Press.

———. 1980. "Cohen on Marx's Theory of History." *Political Studies* 28:121–28.

———. 1982. "Marxism, Functionalism and Game Theory." *Theory and Society* 11:453–82.

Esping-Anderson, Gösta, Roger Friedland, and Erik Olin Wright. 1976. "Modes of Class Struggle and the Capitalist State." *Kapitalistate,* no. 4.

Gorz, André. 1982. *Farewell to the Working Class.* Boston: South End Press.

Jessop, Bob. 1977. "Recent Theories of the Capitalist State." *Cambridge Journal of Economics* 1:353–73.

Wright, Erik Olin. 1983. "Giddens's Critique of Marxism." *New Left Review,* no. 138.

Book Reviews

Capital and the Distribution of Labor Earnings. By Michael Sattinger.
Amsterdam and New York: Elsevier North-Holland, 1980. Pp. xvi +
281. $46.25.

Christopher Winship
Northwestern University

Capital and the Distribution of Labor Earnings is important because work
in economics has not focused on the demand side of labor markets. In
addition, though sociologists have recently created a lot of elaborate ty-
pologies, there is, as yet, little true theory. This book should be welcomed
because it presents a theory.

Much recent research in sociology on the determination of earnings
has attempted to replace the individualistic orientation of status attain-
ment research and human capital theory with a focus on the importance
of industrial, occupational, labor market, and firm structure. Michael
Sattinger has a similar objective in this book, though the theory he de-
velops is comfortably within the tradition of neoclassical and classical
economics. The results are interesting and rich. This book is essential
reading for sociologists working in the new "structuralist" tradition of
earnings determination.

Chapter 1 first describes the basic themes of the book and then outlines
the remaining chapters. The major themes are: (1) that the distribution
and structure of capital in an economy are critical to the determination
of earnings; (2) specifically, the distribution and type of capital will de-
termine the demand for workers of differing ability (defined generally);
(3) as a result, the distribution of earnings will be determined by which
workers are matched to which jobs—which will be a function of the
supply and demand for workers of various types; (4) as a consequence,
shifts in the demand or supply of workers will affect the rates of return
to different characteristics; (5) following from this, there should be a close
relationship between capital accumulation, technological and educational
change, and earnings inequality. If all of this sounds like fairly standard
economics, it is. However, the importance of the book is that Sattinger
develops some specific models from which new insights are gained and
general predictions can be made.

Chapter 2 critiques human capital theory, arguing primarily that it does
not provide an adequate theory of earnings determination since it provides
no theory of demand. The chapter is particularly useful in showing that
there are basically two different models, one developed by Jacob Mincer
and the other by Gary S. Becker and Barry R. Chiswick, that are based

on very different assumptions. However, the argument in this chapter—that the Becker-Chiswick variant needs the "efficiency units" assumption (the idea that individuals differ uniformly in their efficiency across jobs)—is not correct.

Chapter 3 reviews previous work in economics that has focused on how individuals and jobs are matched and the implications for the distribution of earnings. The work that has been done is quite limited. Sociologists should find A. D. Roy's classic article ("Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers* 3 [1951]: 135–46) of considerable interest since it provides a good introduction to Sattinger's book and is quite accessible because it contains no math.

Chapters 4 through 6 provide the theoretical core of the book. Chapter 4 develops an aggregate-level model that relates earnings inequality to inequality in abilities in terms of a function which is dependent on the level of capital intensity and the average level of productive ability (read education) of workers in the economy. Sattinger investigates in detail the conditions under which increasing capital intensity leads to increasing earnings inequality and increasing productive ability leads to decreasing earnings inequality.

Chapter 5 develops a model of matching. Sattinger's interest here is in understanding the conditions under which more productive (higher-ability) workers will be matched with more capital-intensive (expensive) machines. He considers two principles of matching—what he calls the scale of resources effect and the principle of comparative advantage.

Chapter 6 derives a theory of capital costs and earnings from the matching theory. The objective is to show that the assignment principles derived in chapter 5 are consistent with a theory of wages and capital rents and to demonstrate how, as a result, the distribution of earnings will be a function of the distribution of capital.

Many sociologists will probably find chapters 4 through 6 somewhat difficult. Although the math is not particularly sophisticated (many 2 × 2 tables), the economics is technical and the discussion is sometimes hard to follow. Readers may first want to look at some of Sattinger's earlier papers that are cited in the bibliography and at Gary Becker's discussion of the assignment problem (*Treatise on the Family* [Cambridge, Mass.: Harvard University Press, 1981]) in terms of marriage markets. For those readers willing to spend the time to understand the models, the rewards are considerable.

Chapters 7 through 13 present the empirical analysis. All of the analyses pertain to the United States and are quite aggregate in nature. Sattinger's attempt here is not to develop a rigorous econometric specification of a structural model of the theory presented in the previous chapters. Rather, he attempts to test a number of simple hypotheses about the relationship between capital and earnings, using primarily ordinary least squares. The hypotheses that he tests are that: (1) greater capital intensity leads to greater earnings inequality; (2) greater capital concentration leads to greater earnings inequality; (3) increases in average education lead to decreased

inequality; (4) more productive workers will be assigned to more capital-intensive jobs. Using time-series and census data, he finds support for all these hypotheses except the first. From this, two points seem of interest to sociologists. First, the distribution of capital does seem to make a difference in the distribution of earnings. Second, increases in education in the United States have led to decreases in inequality. As Sattinger points out, this finding is in direct contrast to the findings of Jencks et al. (*Inequality* [New York: Basic, 1972]). He argues, and to a degree substantiates, that the effect of increases in education has in good part been offset by increases in capital concentration that have worked to increase earnings inequality.

This book also provides a solution to one of the major problems in the new "structuralist" work in sociology as well as the Marxist economic work from which this sociology has drawn: Why is it that different individuals are found in different sectors of the labor market? I suspect that many sociologists will not be satisfied with Sattinger's answer. It is, however, a start, and perhaps a more interesting economic theory for sociologists to debate than the continued attempts to point out the limitations of human capital theory.

As with a great deal of economics, sociologists will find much to object to in this book. In particular, sociologists will probably object not only to the idea that assignments take place in terms of some optimizing principle; they will also want to reject the unidimensional and abstract way in which capital is conceptualized throughout the book. In a sense there is a contradiction in the book. In the theoretical sections, capital is treated in an abstract and uniform manner. Capital is measured solely in terms of its expense, and results are derived primarily from the assumption that more productive workers are assigned to more expensive machines. Throughout the formal analysis, the point is continually made that the relationship between capital and earnings will be dependent on the nature of the production process. In the empirical analysis, though, capital continues to be dealt with in a unidimensional manner in that differences across the population in the production process (besides in the intensity of capital) are not taken into account. Given the sensitivity of results for the nature of the production process assumed and the fact that few researchers would be willing to accept that the production processes are similar across an actual economy, this approach does not seem reasonable. What, then, are we to make of the empirical analysis? Certainly, the relationships that are being estimated are not structural. Perhaps they are averages of a sort across different processes. But Sattinger has not given us a theory in which averages have any role. Thus, the only proposition validated is that the distribution of capital does affect the distribution of earnings. An important point, but probably not worth seven chapters of empirical analysis.

Let me, however, not end this review on a negative note. This is an important book and has much to contribute to sociologists' thinking about

the demand side of labor markets. The perspective is unabashedly purely economic, but I contend there is much to be said for sociologists' turning some of their attention and debate to good economic theory. The fruit of interdisciplinary dialogue in this area should not be overlooked.

Sociological Perspectives on Labor Markets. Edited by Ivar Berg. New York: Academic Press, 1981. Pp. xvii + 374. $31.00.

Marta Tienda
University of Wisconsin—Madison

Ivar Berg has made an important contribution to the field of labor market studies by drawing together the collection of essays and empirical analyses in *Sociological Perspectives on Labor Markets*. Two features distinguish this set of papers from other volumes in the same genre: first, its emphasis on *sociological* concerns as complements to and extensions of themes formerly dominated by economists; second, its emphasis on *demand* factors that determine individual labor market outcomes. The latter feature makes the book particularly attractive in a field governed by supply-side concerns, especially as more labor market analysts are paying attention to structural dimensions of labor market processes and outcomes. Several essays help clarify conceptual and operational ambiguities surrounding key concepts that have been distorted and misused in recent years, others attempt to elaborate theoretical concerns not well articulated in the past, while others deal in more or less conventional ways with such typical concerns of labor market analysts as unemployment and discrimination. The content of particular chapters ranges from narrowly focused empirical analyses to critical discussions of existing theoretical perspectives and the limitations of these in advancing the current state of knowledge.

Berg's volume grew out of papers presented at sessions of the 1979 meetings of the American Sociological Association and the 1980 meetings of the Southern Sociological Association, both of which he chaired. Teresa Sullivan, who served as discussant for both of these sessions, converted her reflections about the papers into one of the concluding chapters, while Berg seized this opportunity to close the book with his broad research agenda for addressing issues about theories of structural unemployment. Because most of the chapters are prepared by a "new generation of sociologists" whose training in labor market studies has been enhanced by advances in methods and the resurgence of concern for structural determinants of individual outcomes, the book is successful in providing critical insights and, at times, novel ideas. Most of the authors emphasize the-oretical concerns and are "up front" about the value of the use of classical sociological theory for contemporary issues. With one or two exceptions, in other words, there is no pretense of discovering the wheel but, rather, of recalibrating its design to accommodate new advances in the field.

Berg organizes these papers into five sections, four focusing on broad substantive or theoretical concerns, and the closing section discussed above. The first section, which deals with theoretical perspectives of matching persons to jobs, includes one chapter by Mark Granovetter and another by Aage Sørensen and Arne Kalleberg. Both chapters focus on how the different intellectual backgrounds of sociology and economics can be mutually complementary and lead to critical insights about the operation of labor markets, but Granovetter's piece does this more methodically. Both papers demonstrate that sociologists have a decent command of the economic literature, one which is not reciprocated by economists, and both highlight the promise of a "matching" approach for understanding labor market dynamics. However, Sørensen and Kalleberg elaborate more fully the conceptual detail of this approach and its application. Unfortunately, their specifications for testing the framework empirically are not as fully developed as their ideas, and they close lamenting that an excellent opportunity to test the framework was missed.

The second major segment of this book deals with the structural dimensions of the labor market. This section also contains two chapters (both coauthored by Kalleberg) and emphasizes the importance of economic organization for understanding unequal labor market outcomes. Although numerous labor market analysts have reached this same conclusion in recent years, the paper coauthored by Kalleberg and Michael Wallace stresses the need to differentiate between economic organization of firms and industries and economic segmentation as these dimensions relate to the problem of control. Of course, this distinction is not as easy to manage operationally as theoretically, and the empirical analyses in this section do not resolve the problem. Nevertheless, both chapters in this section are useful for their systematic distilling of master labor market concepts that have been misused and operationally misrepresented in several recent attempts to evaluate empirically how labor market organization and structure determine worker outcomes. In their chapter, Robert Althuser and Kalleberg properly call for a more thoughtful treatment of labor markets, including a proper differentiation of internal, primary, and occupational labor markets and a distinction between labor markets and economic sectors.

The two sections about unemployment and labor market discrimination differ from the preceding two in that most chapters (three per section) are empirical analyses focused on specific questions. Paul Schervish's analysis of unemployment confirms an unsurprising finding—that the structure of unemployment is related to the structure of employment— but he deserves credit for his painstaking efforts in theorizing which of the various aspects of labor demand should influence different types of unemployment (i.e., short layoffs, firings, long layoffs, and quits) and how they influence them. One could quibble about whether log-linear models permit him to test his hypotheses adequately, but an even more disturbing problem is that he succumbs to the strategy of using a single

indicator to represent the structural dimensions of labor markets (oligopoly, competitive, farm, and construction sectors). This is in spite of the conclusions of the previous chapters that market structure is far too complex to be represented in a single dimension.

Daniel Cornfield's empirical analysis of unemployment tests a series of hypotheses about how downturns in demand affect variation in layoff rates among manufacturing firms. Cornfield's results show two factors—job structure and expectation of recall—to be especially strong determinants of variation in manufacturing layoff rates. What is disappointing about this analysis is that his measure of job structure, the key to his results, does not live up to its theoretical development and, consequently, one is left wondering what the results actually mean. Toby Parcel's chapter, "The Development and Functioning of the American Urban Export Sector, 1947–1972," while a competently executed piece, does not fit as neatly into either this section or the entire collection of papers.

The chapters about discrimination are least innovative in the questions they pose and the methods used to address them. Granovetter would classify the paper by Alice Kemp and E. M. Beck about "Female Underemployment in Urban Labor Markets" as a revisionist version of the status attainment and human capital research traditions which give explicit causal priority to individual-level variables even while acknowledging the importance of structural factors. I am equally disappointed by the absence of solid explanations of how and why the potentially interesting aggregate indicators used in the analyses operate as they do. Finally, Robert Miller's piece attempts to determine empirically whether local opportunity or cultural heritage was more important in determining the difficulties European immigrants experienced in securing employment during the Great Depression. Since this is an almost rhetorical question these days, his conclusions are not surprising. Unfortunately his data do not permit an adequate test of this question. Miller deserves credit for his systematic culling of testable hypotheses about immigrant labor market success from assimilation theory. Some of these already have been addressed more rigorously by contemporary immigration analysts, but others must wait for more appropriate data to determine exactly how variations in local opportunities influence the labor market integration processes of immigrants.

In short, I believe this collection of readings represents a solid contribution in its own right, but those papers that strive for conceptual advancement and synthesis of existing ideas go furthest toward illustrating the sociological dimensions of labor market processes and outcomes. These chapters are accompanied by extensive bibliographies and literature summaries that should be of interest to individuals beginning their training in labor market analysis or those who are attempting to weave the various pieces of research claiming to account for demand factors into a coherent framework. What is clear, after all is said and done, is that we still have a long way to go.

Segmented Work, Divided Workers: The Historical Transformation of Labor in the United States. By David Gordon, Richard Edwards, and Michael Reich. Cambridge: Cambridge University Press, 1982. Pp. xii + 288. $34.50 (cloth); $9.95 (paper).

Eric L. Hirsch
Columbia University

A group of economists associated with the Union for Radical Political Economics have been engaged in research that has been of great interest to sociologists since the 1960s. This latest effort by David Gordon, Michael Reich, and Richard Edwards, entitled *Segmented Work, Divided Workers: The Historical Transformation of Labor in the United States,* analyzes the structure of the working class and the labor market and describes management techniques that have been used to discipline the labor force since the rise of capitalism in the United States. The authors reject the popular Marxist approach which divides U.S. economic development into two stages—competitive and monopoly capitalism. Instead they propose that the system has progressed through three stages, with each stage representing a unique solution to the problem of how to accumulate capital. A *social structure of accumulation* is created in the *exploration* phase of each stage; that structure consists of all the institutions that impinge on the capital accumulation process: investment institutions that provide money and credit; political institutions that provide public services, enforce regulations, and create tax burdens; and labor unions that limit corporate power. In the *consolidation* phase of each accumulation stage, the social structure of accumulation allows a period of great economic expansion. But eventually, internal contradictions lead to a period of economic crises, a *decay* phase during which the next social structure of accumulation is explored. This new structure then allows another period of economic expansion until it, in turn, is undermined by contradictions. These cycles of expansion and contraction are not identical with the business cycle; there are generally several business cycles within each stage, within each *long swing* of capital accumulation.

The first accumulation stage in the United States—*proletarianization*—occurred in the 1820s to the 1840s. This stage was characterized by the growth of a wage labor force divorced from the means of production. Many formerly independent producers—agricultural workers, artisans, and also women and children—were used as sources of labor. Immigration was crucial in the consolidation phase of proletarianization which began in the 1840s and continued until the 1873 depression. Problems of competition, between capitalists and the continued ability of craft unions to control shop-floor production, led to a profit squeeze and to economic crises in the decay phase from the 1870s to the 1890s.

The capitalist response was the creation of a new social structure of accumulation, the *homogenization* stage, explored in the late 19th century and characterized by mechanized production and more direct supervision

by foremen. This *drive system* of production resulted in increased plant sizes, growth in the number of semiskilled operatives, and a narrowing of the economic and social gap between the skilled and unskilled within the working class. Consolidation through the World War I period was facilitated by the turn-of-the-century merger movement that produced giant corporations with the ability to pursue homogenization more aggressively. Contradictions within this system—especially high labor turnover, informal output restriction agreements by production workers, and lack of effective demand owing to low wages—led to crises beginning in the 1930s Depression.

The third stage, *segmentation,* was explored in the period between the wars and consolidated in the post–World War II years. It involved the explicit recognition of workers' power in industrial and craft unions and led to the creation of a *dual labor market*. Large firms built internal labor markets with unionized jobs, high worker wages and benefits, promotion from within the firm, and bureaucratic systems of control; smaller firms remained nonunion and provided only low wages, no benefits, and dead-end, insecure jobs. Previous patterns of race and sex discrimination were reinforced by the dual labor market as less privileged groups found it difficult or impossible to gain entry to the better, primary sector jobs. This segmentation stage is now in decay, as reflected by the economic recessions of the mid-1970s and early 1980s and by the refusal of large corporations to honor the terms of the post–World War II accord with the major labor unions.

There are several problems with the authors' analysis. They have difficulty in marshaling support for their second, homogenization stage. The expansionary part of the stage did not show dramatic GNP growth as did the other two stages, and there was no decline in wage differentials between the skilled and unskilled. Ethnic, race, and sex fragmentation within the working class continued unabated in the early part of the 20th century, reinforced by employer divide-and-rule tactics. Finally, until the 1930s the unskilled and semiskilled rarely joined the skilled in union organizing. The reality in this period seemed to be continued economic and social fragmentation of the working class and not homogenization based on skill degradation in the skilled trades. These problems cast doubt on the utility of a three-stage versus a two-stage theory of U.S. capitalist development.

Also, Gordon, Edwards, and Reich never consider the question of what might be beyond the segmentation stage. There is no detailed discussion of the reasons for the decay of that stage or the possible nature of stage 4. And they do not address the issue of the long-term survival of capitalism. Will it persist indefinitely, based on the recreation of new social structures of accumulation? Or will it run out of options consistent with the private enterprise system? Actually, the fact that such questions can be addressed to the authors at all reflects the power of their analysis. The book successfully combines detailed, comprehensive historical description with a theoretical framework based on the long-swing social structure of

467

American Journal of Sociology

accumulation connection. A wide variety of disparate historical materials are thus brought to bear on the problem of the capitalist-worker relationship. It is the authors' method that is most exciting here; it is left to sociologists, labor historians, and economists to realize the potential inherent in that method.

A Treatise on the Family. By Gary S. Becker. Cambridge, Mass.: Harvard University Press, 1981. Pp. xvi + 288. $20.00.

Arthur L. Stinchcombe
University of Arizona and *Stanford Graduate School of Business*

Sociologists will have to get into a special mood to read this book, because, by sociological standards, its style of reasoning is outrageous. Since I believe many of the ideas in *A Treatise on the Family* will prove valuable in family sociology, I will try to harden the prospective sociological reader who would otherwise be driven up the wall.

First, the book is insistently mathematical. The mathematics is all accessible with first-semester calculus—there are no integral signs—plus some experience in reading economics. The habit of thinking of curves with elasticities, or the immediate recognition of what a Cobb-Douglas production function looks like, is necessary to read the mathematical appendixes to chapters. There is nothing in the mathematics that cannot easily be explained in words, and with work one could simply ignore the mathematics. The reasoning is, however, much denser than we are used to seeing, and one must read more slowly.

Second, Gary Becker likes to reason from outrageous assumptions. Chapter 2 is based on the premise that a couple produces sexual satisfaction or children or other utilities jointly (true enough), then they divide the satisfaction in accordance with the marginal contribution of each; the day when your share is two-thirds of an orgasm and the left leg of a cranky child, you know your contribution was deficient. It is assumed, at least from page 116 to page 145, that the better endowed a child is to make money as an adult (e.g., the more a child is a bright, white male), the less parents will invest in the child's health and education. In general, what parents are assumed to want out of a child is the child's adult income (see esp. p. 114), so, for example, if parents see that their children will be taxed heavily, they will educate them less or care for their health less (see esp. chaps. 6 and 7). On page 156 the extreme proposition is derived that anticipations of economic development, by raising expected incomes of children regardless of investment, will decrease investments in children's health and education.

Third, the outrageous assumptions are embedded in interpretations of facts. On page 56 the assumption that wives get an "income" of satisfactions from their marriage equal to their value on the marriage market

468

is interpreted to imply that, in societies with bride-price (showing that the bride is more valuable), the bride-price is paid to the bride. On page 111 the fact that mortality of children is higher among the poor than among the rich is interpreted to mean that parents can control the mortality of their children by investing more in them. On page 71 assortative mating by biological traits such as height (the correlation of heights of spouses is about equal to the correlation of IQs of spouses) is interpreted in terms of the marginal productivity of the trait—Becker does not say whether taller or shorter is better for producing the satisfactions of marriage. On page 235 he estimates the divorce rate of previously divorced or widowed women to be 2.8% per year higher during the first five years of the second marriage than that of otherwise equivalent women in first marriages. Becker interprets this as owing to the greater quarrelsomeness of the average divorced person. Aside from the implication that the widows must have been quarrelsome enough to drive their former husbands to their graves, this means that divorced men are only a quarter as quarrelsome, for they have a rate only 0.7% per year higher than those in first marriages. Here the overall R^2 in the equation for newly (re)married women (the largest R^2) for all variables Becker specifies as relevant is under 4%, which does not bode well for the explanatory power of economic theory in this area.

Fourth, part of the derivation of empirical predictions well-known to be false (this happens throughout the book) comes from using simplifying assumptions which can later be dropped. For example, in chapter 3, on "Polygamy and Monogamy," we start the analysis of polygamy (p. 47) with the assumption that men differ considerably in "efficiency" (on p. 58, we find out that "abilities" that enter into "efficiency" include inherited capital), while women are identical. Efficient males' creation of excess demand for women raises the equilibrium price of women. Since we have already assumed that women get their equilibrium price, women are therefore shown to be better off in polygamous societies. It is here that the mistake about the bride-price is relevant. Since usually the male guardians of a young woman get the bride-price, the fact that polygynous societies are more likely to have bride-prices shows that in polygynous societies the difference between the marginal productivity of a wife and what the husband provides as her income is increased—that is, it implies exactly the opposite conclusion from Becker's—that women are worse off in polygynous societies. Becker needs to read Karen Paige and Jeffery Paige's *Politics of Reproductive Ritual* (Berkeley: University of California Press, 1981). At any rate, later the sillier derivations are seen to have been products of the oversimplification that women are all alike and can be dropped from the theory.

Why, then, should sociologists read the book? It seems to me that many of the models are onto something. They are badly formulated and carelessly tested, but the kernel of the right theory (or of part of the right theory) is there. If a more sophisticated theory can capture the kernel, and yet preserve some of the simplicity of the theoretical structure, it will

469

be to the benefit of all of us. Becker seems, perhaps, a little simpleminded, but that should not blind us to the fact that simplicity is a virtue, and even worth some sacrifice of realism: not this much sacrifice, but some.

Guardians of the Flutes: Idioms of Masculinity. By Gilbert H. Herdt. New York: McGraw-Hill Book Co., 1981. Pp. xviii+382. $17.95.

D. S. Gardner
Australian National University

Guardians of the Flutes is at once an intriguing and important contribution to ethnographic knowledge and sad testimony to the theoretical fragmentation of the sociological/anthropological sciences. It is primarily concerned with the patterns of institutionalized homosexuality and familial heterosexuality among the pseudonymous "Sambia" of Papua New Guinea. The subject matter of the book will, inevitably, ensure the interest of those in a number of disciplines and probably that of many nonspecialists, too.

Gilbert H. Herdt sets out to describe and analyze the impact of daily homosexual practices (which, somewhat tendentiously, he refers to as "ritualized" fellatio) upon the sex identity of Sambia males. Between the ages of seven and 10 years Sambia boys are initiated into the performance of fellatio on bachelors. The explicit aim is to enable them to grow up to be adequate men. Later these fellators assume the role of inseminators for their own ritual juniors. However, once men have begun to indulge in heterosexual intercourse with their wives, then they should cease to have any further homosexual encounters. These practices are accounted for by the Sambia in terms of the beliefs and values of an exclusively male secret cult which is dominated by senior men. This cult, we are told, is a particularly important factor in the inculcation of the values and attitudes of warriorhood. The viability of community life prior to European contact was dependent on having a force of fierce warriors, since warfare was endemic.

These martial values are an integral part of a wider view of the relations between the sexes, the nature of femininity and masculinity, and the principles underlying them. Although the most secret myth associated with the cult (imparted to men only after they have become fathers) might suggest otherwise, ritual dogma is evidently predicated on the belief that females mature naturally and more quickly than males, who attain manhood only with the aid of ritual techniques and the ingestion of semen.

One of the merits of this bold work is the attention the author pays to the developmental contexts of these practices and beliefs and to their consequences for the psychological constitution of Sambia men. Unfortunately, this important project, which embodies research interests unusual among anthropological fieldworkers, is less impressive than it might have been. This is because of Herdt's theoretical posturing and his ten-

470

dency toward rhetorical excess. Durkheim, in particular, as the putative
founding ancestor of the social anthropological clan Herdt finds so dis-
agreeable, comes in for harsh treatment. However, Durkheim is, I submit,
unrecognizable in the flimsy straw man Herdt sets up as a target. For the
most part the author's theoretical antics are merely irritating in their
speciousness; however, at certain points they become plain silly.

On the dust jacket we are told that Herdt's remarkable findings are
drawn ". . . from careful observation and from the *actual words* of the
participants" (my emphasis). While there is no doubt that many of the
data are remarkable, it is entirely unreasonable to suggest, as Herdt does,
that talking to people and observing their behavior is unusual in social
anthropological research. The point Herdt really wants to make is that
social anthropologists (who he seems to think, as he indicates on page 5,
form a single theoretical school) have missed much in ignoring the specific
qualities of the experiences particular individuals have in various socially
prepotent settings. But in his eagerness to further his claim that his
approach is innovative, he resorts to hyperbole of the magnitude exem-
plified in the following: "This book stems from a premise that anthro-
pologists usually ignore individual verbal behaviour and experience for
understanding identity, communicative acts, and ritual symbolism" (p.
58); ". . . the . . . Durkheimian viewpoint [is that] an observer, any ob-
server, simply by watching a ritual, is able to infer the shared norms and
meanings of its participants" (p. 10). Even Radcliffe-Brown, who is surely
an arch-villain by Herdt's lights, made it perfectly clear that the only
concrete data available to the anthropologist, and the only basis for so-
ciological analysis, are precisely the speech, actions, and interactions of
particular individuals. These excesses must create doubts about the care
with which Herdt has constructed his general case in spite of his metic-
ulousness in presenting and analyzing his ethnographic data. (Although
an exception here is his grossly mistaken characterization of the Sambia
kinship terminology as an "Omaha" type.)

My criticisms are not made in a spirit of devotion to the form of social
anthropology which does have its roots in certain aspects of Durkheimian
sociology. On the contrary, there are problems here and I agree with
Herdt that they are most apparent in the orthodox approach to ritual.
Nevertheless, I do not think that Herdt has succeeded in pinpointing
these deficiencies. The strengths of the book are manifest in his ethno-
graphic acuity, but even here the weak theoretical apparatus he constructs
sometimes will not bear the weight he places on it.

"Idioms" is a key term in the author's analytical scheme. These are
understood as ". . . characteristic expressions having peculiarities of cog-
nitive and emotional meaning and style established in . . . ordinary in-
terpersonal communications . . ." (p. 14). He also sees them as expressive
of the ritual traditions experienced by the Sambia who use them. Putting
to one side any reservations about this view and its unusually wide con-
ception of "idiom," consider the case of the cassowary. Sambian idioms
concerning the cassowary's reproductive characteristics are said to be

fantastic. "Fantasy . . . is at play when everyday comments . . . clearly transcend the ethnographer's descriptions of perceptible reality, and when the natives' expectations fail to be restrained by perceptual information but instead refer, not to the sensations of a phenomenon, but to its imagined character" (p. 132). Aside from the fact that this passage seems odd coming from someone who castigates others for their positivism, it turns out that the Sambia are not so far from the mark in saying that the cassowary gives birth through its anus or that it is a masculinized female. Herdt does not mention that the cassowary has a single urogenital opening (the cloaca) from which both feces and eggs are deposited. Nor does he point out that this bird, along with other ratites (emus, ostriches, and kiwis), is unusual in that the male is endowed with an erectile penis. This penis is located inside the cloaca, and apparently the Sambia (pp. 146, 147), in common with other people in New Guinea, believe the cloaca to be similar to a vagina in certain respects. Of itself this evidence does not establish much, but it does suggest that great care is required in sorting out what is reasonable belief and what is fantasy in people whose lives are very different from our own. Elsewhere Herdt makes much of the use of "idioms" in which the word "nose" is used to allude to the penis. The reader is subsequently told, however, that the same lexeme denotes both anatomical organs.

Other important terms used by Herdt in the construction of his theoretical position are also left virtually unanalyzed. His apparent obliviousness to, and confusion of, the various senses of that most treacherous of terms, "meaning," his frequent but unexplicated use of the notion of subjectivity, and the promiscuous use of the concept of a symbol (traits which, ironically enough, are the plague of the anthropological approaches to ritual which he eschews)—all serve to undermine his attempts to master the fundamental problems which, to his credit, Herdt does succeed in raising. However, the concepts he uses are no more up to the task he has set himself than are handyman's tools fit for surgery.

Much more could be said about the theoretical and evidential shortcomings of the book. But since others dealing with the Sambia are to follow this one, many of these further criticisms may be undercut. However, it is to be hoped that the other major flaw of this work will not mar these later volumes. This is the idiosyncratic and often barbarous English in which it is written—something which should be subject to much tighter editorial control in the forthcoming books.

Human Nature and History: A Response to Sociobiology. By Kenneth Bock. New York: Columbia University Press, 1980. Pp. x + 241. $18.95.

Frank Spencer
Queens College CUNY

In 1975, Edward O. Wilson published a book, entitled *Sociobiology: The New Synthesis,* in which he set forth the general biological principles that

(presumably) govern social behavior and social organization in all kinds of animals, including-*Homo sapiens*. Succinctly, Wilson and his followers claim that all behavior can be explained in biological terms, namely, by differential processes of genetic transmission. Had Wilson restricted his geneticism to a consideration of arthropods and colonial jellyfish, I suspect very few social scientists would have been aware of either his work or the so-called sociobiological movement.

While most sociologists and anthropologists are willing to accept that there are biological factors that affect human behavior, they are not prepared to accept the sociobiological proposition that human culture is a genetically predetermined system. According to the sociobiologists, human mating patterns are genetically determined. But, as Marshall D. Sahlins has countered in his book *The Use and Abuse of Biology: An Anthropological Critique of Sociobiology* (Ann Arbor: University of Michigan Press, 1976), ". . . people don't do what their genes tell them; polyandry isn't a genetic process, it's a form of received knowledge or accepted wisdom of the group." Although Wilson and other sociobiologists would contend that reality is otherwise, they have in fact failed to provide a satisfactory account of exactly how their gene-based theories are supposed to work in the *real* world of human affairs.

In seeking to show the autonomy of human sociocultural systems from direct biological causation, and thereby argue that the genus *Homo* is psychobiologically unique in having both an organic evolution and a sociocultural history, Kenneth Bock (a sociologist from the University of California) takes the position in *Human Nature and History: A Response to Sociobiology* that human cultural behavior is significantly shaped by history. Therefore, Bock contends, human nature is more a product than a producer.

As Bock recognizes, the roots of this scientific, but largely philosophical, debate can be traced back both to Aristotle and to the attempts during the 17th and 18th centuries to codify man's place in nature. On the one hand there were those workers who regarded man as simply a fixed organismic machine, while others assumed a more plastic (dynamic) viewpoint.

Following in the wake of Darwin's thesis *The Origin of Species* (1859), a number of social scientists endeavored with varying success to incorporate these diametrically opposed ideas into theoretical schemes that would explain the genesis and development of human society. Among these were Lewis Henry Morgan and Friedrich Engels. Although these workers were attempting to operate in a mechanistic-evolutionary matrix, nevertheless, there is a vitalistic element embedded in their ultimate conclusions. In other quarters, however, this vitalism became the organizing theoretical principle. The reasons for this seemingly paradoxical amalgamation are, of course, complex, but important because they do much to explain not only the subequent intellectual schism between biology and the social sciences at the close of the 19th century (particularly here in the United States) but also the incipient antievolutionary stance of anthropological theory at that time. Unfortunately, however, Bock's treat-

ment of this issue is unsystematic, and as a consequence, readers unfamiliar with the convoluted intellectual history of the social sciences during the latter half of the 19th century will have some difficulty in following the subsequent (differential) fate of Darwinian theory in the hands of biologists, ethnologists, and other social scientists during the 20th century.

I hasten to add, however, despite this criticism, that Bock has succeeded in providing a splendid historical overview of anthropological thought to Darwin. Among the various scholars he discusses are Descartes, Rousseau, Lord Monboddo (James Burnett), and Montaigne, to mention just a few.

The major portion of this book, however, is given over to a careful consideration of the sociobiological synthesis as it has emerged during the past few decades. Beginning in chapter 3, "Sociobiology and the Human Sciences," Bock itemizes the major difficulties attending acceptance and verification of the central thesis of sociobiology, the. formative influence of genes on the production of sociocultural phenomena. These "difficulties" are then subjected to careful examination and elaboration in four subsequent chapters.

Among the "difficulties" raised by Bock is the question of human cultural diversity. The sociobiological response to this problem is to argue that what has evolved, and what is genetically determined, is the inherent flexibility or plasticity of the human biogram. But Bock contends (as indeed Sahlins has in his critique) that this is simply not an adequate explanation. Although human cultures may well be cases of adaptation, the issue is whether that adaptation is genetic or not, and, as Bock points out repeatedly, the sociobiologists have failed to provide us with the proof that cultures are genetically programmed.

In probing the "evidence" (or rather the lack of it) and conclusions of the sociobiologists, Bock reasserts constantly the importance of a historical perspective in understanding the novelty of the human condition. Indeed, it occurred to me while reading this book that a probable reason why so many workers in the biological sciences have been led to embrace (so uncritically) the distorted dogma of sociobiology is because of the general lack of a historical perspective in their intellectual preparation. In contrast, social scientists are, by definition, historically oriented. As such, they are acutely aware of their own intellectual history and the earlier attempts by social theorists to search for such entities as a "Geist," "élan vital," "cultural spirit," or "racial genius" lying behind and giving shape to human histories. Such quests in the past, have, Bock notes: ". . . led students of human social and cultural phenomena to lose the empirical-historical referent of their propositions. . . . It is the most primitive and simplistic mode of explanation in man's contemplation of the world about him. There should be understanding, at least, of humanists' wariness of the most recent suggestion that genes are to be regarded as the basic architects of history" (pp. 160–61).

This book is not, and was not meant to be, a definitive account of the sociobiological debate. However, it does present a well-reasoned and

convincing rebuttal of the sociobiological thesis, thereby demonstrating from a fresh perspective the proposition that sociocultural phenomena must be explained and understood in their own terms.

I recommend this well-written book highly and intend to use it in my undergraduate course on the history of anthropological theory.

Sociobiology Examined. Edited by Ashley Montagu. New York: Oxford University Press, 1980. Pp. x + 355. $19.95 (cloth); $5.95 (paper).

Mark V. Flinn
Northwestern University

During the 1960s and early 1970s, a number of important theoretical developments occurred in evolutionary biology. New theories of nepotism, reciprocity, parent-offspring interactions, and mating systems emerged. These developments were sparked by the recognition that natural selection operates primarily at or below the level of the individual (G. C. Williams, *Adaptation and Natural Selection* [Princeton, N.J.: Princeton University Press, 1966]). Adaptations previously thought to have evolved for the benefit of the group or species (sterile workers in insect colonies, or alarm calls, e.g.) are now explained in terms of individual advantage. The publication of E. O. Wilson's *Sociobiolgoy* (Cambridge, Mass.: Belknap, 1975) attracted considerable attention to the new developments in biological theory and their relevance for understanding human social behavior.

Unfortunately, the controversy that ensued was often emotional and uninformed. Many of the essays in *Sociobiology Examined,* edited by Ashley Montagu, offer such polemics. Some of the contributions—Montagu) (chap. 1), M. Midgely (chaps. 2, 5), and S. Rose (chap. 7)—are mired in character assassination and bad politics. Most of the papers show little understanding of modern evolutionary theory—J. Barkow (chap. 8) is a notable exception—and instead chase the phantom of genetic determinism.

Part of the blame for this reactionary dismissal of sociobiology can be laid at Wilson's feet. The use of biological terminologies, the undocumented speculations, and the inadequate analysis of the proximate mechanisms of cultural transmission in *Sociobiology* and *On Human Nature* (Cambridge, Mass.: Belknap, 1978) alienated many social scientists. S. J. Gould (chap. 12) and N. J. Macintosh (chap. 15) criticize these and other shortcomings of Wilson's discussions of human behavior but fail to address the strengths of the evolutionary theories reviewed by Wilson.

Indeed, most of the essays in *Sociobiology Examined* are plagued by the old culture-biology, nature-nurture bugaboos. Montagu (chap. 1), K. Peter and N. Petrysak (chap. 3), S. A. Barnett (chap. 6), Rose (chap. 7), D. Freeman (chap. 9), S. Washburn (chap. 11), and M. Harris (chap. 14)

dismiss sociobiology with the arguments that (a) culture change can occur without concomitant genetic change, and (b) cultural variation can occur in the apparent absence of correlative genetic differences. To my knowledge, no sociobiologist has suggested otherwise.

The confusion surrounding the relationship between culture and biology stems from the fact that cultural transmission occurs via social learning, independent of genetic transmission per se. This has led many students of human social behavior to conclude that culture is a phenomenon sui generis, decoupled from organic evolution. As Washburn states, "the laws of learning are not the laws of genetics" (p. 278). Both sets of "laws," however, have evolved by natural selection—unless we are to invoke divine creation.

Learning capabilities and constraints would not have evolved (i.e., been favored by natural selection) if they were random with respect to biological adaptation. We have evolved to learn adaptively and have *evolved* to learn nothing else. This is not to deny that we commonly and frequently learn maladaptively; imperfection is the bedfellow of unpredictable environments. Hence, perhaps, our evolved abilities to observe and learn from the mistakes and successes of others. But to claim that learning decouples culture from biology is to ignore the cumulative effects of a history of natural selection (see R. D. Alexander, *Darwinism and Human Affairs* [Seattle: University of Washington Press, 1979]).

The question posed by sociobiology is *not* to what extent is social behavior "genetically determined" but, rather, Is culture adaptive in the *new,* biological sense of that term? Montagu's volume attacks the straw men in the first question. The relevant question is addressed by such recent works as Napoleon Chagnon's studies of the Yanomamo (see Chagnon and Irons, *Evolutionary Biology and Human Social Behavior* [North Scituate, Mass.: Duxbury, 1979]), and Martin Daly and Margo Wilson's studies of child abuse ("Abuse and Neglect of Children in Evolutionary Perspective, pp. 405–16 in *Natural Selection and Social Behavior,* ed. R. D. Alexander and D. W. Tinkle [New York: Chiron, 1981]) and homicide ("Homicide and Kinship," *American Anthropologist,* in press).

Sociobiology Examined does not critique empirical studies. It is full of basic misunderstandings of biological theory and misleading snipes at Wilson's speculations. I recommend it only to those interested in the history of the sociobiology debate.

The Philosophy of Evolution. Edited by Uffe J. Jensen and Rom Harré. New York: St. Martin's Press, 1981. Pp. vii + 299. $22.50.

Neil M. Ribe
Lamont-Doherty Geological Observatory of Columbia University

Argument by analogy is a time-honored method of reasoning but is not without its dangers. Analogy, by proceeding from the well-known to

something less known, often gives unique insight into questions otherwise difficult to approach. At the same time, however, analogy can easily be overextended to give the appearance of explanation where true understanding is lacking.

The use and abuse of analogical reasoning are especially apparent in modern understandings of "evolution" which borrow concepts from biology and apply them to the social order. The validity of such understandings is the subject of *The Philosophy of Evolution*. The book aims, as its dust jacket states, to provide "a systematic analysis of the wide variety of contexts" in which evolutionary concepts have been applied, but actually addresses only three: biological evolution (pt. 1), the development of social structures (pts. 2, 3), and the growth of knowledge (i.e., science; pt. 4).

Any discussion of an analogical concept (such as "evolution") should, before all else, provide a thorough understanding of the concept in its original field of application. This *The Philosophy of Evolution* largely fails to do. The responsibility for providing an overview seems to rest with the editors, Uffe J. Jensen and Rom Harré, but Jensen's introduction to the volume is poorly written and decidedly unhelpful. The two papers that make up part 1, however, address substantive issues. D. Hull's "Units of Evolution: a Metaphysical Essay" seeks to answer the question of *what* actually evolves by replacing the usual range of biological answers—gene, organism, species, population—by more general notions of "interactor," "replicator," and "lineage." The evident applicability of such abstract concepts to social and cultural development, however, may have been bought at a high price: it is not obvious that they increase our understanding of *biological* evolution. P. Beurton's "Organismic Evolution and Subject-Object Dialectics" suggests that organisms are not simply passive objects at the mercy of their environment but that, through behavior, they have an active role in shaping their own evolution. Beurton's notion has at least the virtue of suggesting an alternative analogy, one less abstract than Hull's, between biological evolution (organism as agent) and cultural evolution (man as agent). Beurton's essay, however, which stands outside the main neo-Darwinian tradition of modern biology, shares to some extent a major weakness of *The Philosophy of Evolution:* there is little discussion of the doubts and qualifications concerning the theory of natural selection expressed by numerous biologists and philosophers of science during the last century. The recent ferment over the notion of "punctuated equilibrium," for example, is not even mentioned.

In parts 2 and 3, which deal with the social and cultural applications of evolutionary concepts, two essays are notable for their balanced appraisal of the merits and limitations of evolutionary analogy. R. Döbert's "The Role of Stage Models within a Theory of Social Evolution" raises the crucial question of mechanism: If the mechanisms of biological and cultural evolution are essentially different, how can the former be used to illuminate the latter? Insofar as stage models of social change fail to specify the mechanism of change, in Döbert's view, they are essentially

descriptive rather than explanatory. One of the better essays in the volume is S. Toulmin's "Human Adaptation," a discussion of four "modes" of adaptation: "calculative" (conscious choice), "homeostatic" (regulatory mechanisms), "developmental" (morphogenesis, psychological maturation), and "populational" (natural selection). The *dominant* adaptational modes in the biological and sociocultural spheres may be different, Toulmin notes, but, in general, several or all of these modes will be operative in each case.

Part 4, "Approaches to the Growth of Knowledge," is largely devoted to a debate over the validity of realist interpretations of scientific theories. While the essays are worthwhile, they leave the question of whether science is "evolutionary" largely unanswered, and hence will appear somewhat tangential to the central purpose of the book.

This volume is based on a symposium held at the Institute of Philosophy at the University of Aarhus, Denmark; several of the participants will probably be unfamiliar to American readers. Regrettably, the same must be said of much of their prose—English is not the native language of several of the contributors, which makes for difficult going at times. *The Philosophy of Evolution* is an uneven book, and leaves many important issues untouched. Nonetheless, it contains a sufficient number of substantial essays to make selective reading worthwhile.

The Social Basis of Scientific Discoveries. By Augustine Brannigan. New York: Cambridge University Press, 1981. Pp. xi + 212. $24.95 (cloth); $9.50 (paper).

Naomi Aronson
Northwestern University

Both the folklore and the scholarly literature on creativity in science and the arts make much of the flash of insight, the "aha" sensation of the Gestalt shift that marks the genius at work. Archimedes discovered the laws of hydrostatic displacement in the bathtub; Newton discovered the law of gravity in a falling apple. All such anecdotes celebrate our amazement at having pointed out to us what had been in front of our noses all along. Augustine Brannigan's *The Social Basis of Scientific Discoveries* is, as Robert Merton is so fond of saying, a self-exemplifying case. The first half of the book left me thinking, "Of course, it's so obvious! Now why didn't I think of that?"

Brannigan reviews the literature on scientific discovery and concludes that everyone has approached the problem backward. Philosophers, historians, and sociologists, starting with examples of discoveries, have tried to explain how these discoveries occurred. Most accounts of discovery are psychological; they explain how individuals get innovative ideas. Critics of such mentalistic models point out that the same innovative idea

478

often occurs to several individuals independently and simultaneously ("multiple discovery"); therefore, cultural accounts suggest that scientific discoveries occur because somehow the "time is ripe." Both types of explanations err, Brannigan says, in assuming that the status of discovery is an unproblematic one; what we really need to know is how scientists define something as a discovery in the first place. Brannigan proposes an attributional model of scientific discovery in order to "explain how certain achievements in science are *constituted* as discoveries . . ." (p. 11).

Sociologists, who, after all, pride themselves on transcending psychological explanations of phenomena, will be surprised to learn that the presuppositions of the mentalistic model have penetrated recent sociological thinking about science. The culprit, it seems, is Thomas Kuhn, whose *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1970) has influenced sociological thinking not just about science but also about all kinds of cultural and intellectual change. Brannigan notes that there are two distinct sociological aspects in Kuhn's work: "an account of how discoveries are made by individuals, and an account of how communities resist or accept such discoveries" (p. 21). For Kuhn, the perception of anomaly by individual scientists is the necessary but not sufficient condition for scientific revolutions. Brannigan shows how, by attributing the individual scientist's perception of anomaly to the process of Gestalt switch, Kuhn's theory incorporates the weaknesses of the mentalistic model.

The term "Gestalt switch" refers to the perceptual experience in which an entire pattern of observations and their presumed interrelationship is rearranged. Brannigan suggests that, in focusing on the Gestalt switch, the mentalistic model has mistaken the cognitive and affective experiences that accompany discovery for the cause of discovery. But the Gestalt switch is not unique to scientific discovery. In other contexts the phenomenon is called "insight" or "learning." Students performing textbook exercises, clients in psychotherapy, and scientists whose innovative ideas are erroneous, all experience Gestalt shifts. Indeed, Brannigan observes dryly, mentalistic explanations of scientific discovery distinguish between correct and erroneous insights only by attributing the former to cognitive processes and the latter to psychopathology.

Brannigan shows that the cultural explanations of discovery that counter the mentalistic model are equally unsuccessful. One important criticism is that cultural explanations cannot account for changes in the social status of a scientific achievement. For example, palaeoanthropologists considered the Piltdown man to be a valid discovery of the missing link between the higher apes and *Homo sapiens* for almost 40 years before dismissing it as incorrect and probably fraudulent. An adequate model of discovery must show both how the discovery was accepted as valid and how it was rejected.

Brannigan's attributional model of "discovery as meaningful action" identifies four criteria by which individual scientists and the scientific community define scientific achievements as discovery. First, the an-

nouncement of discovery must be congruent with the expectations of the scientific community; it must fit with what is already known in the field or resolve outstanding problems. Second, scientists must perceive the announcement as resulting from legitimate scientific inquiry, not from other activities, such as fraud or alchemy. Third, the announcement must be perceived as correct. Fourth, the announcement must be perceived as unprecedented. This last criterion, of course, is the origin of priority disputes; only the first scientist to announce a particular finding gets the credit for discovery.

The Social Basis of Scientific Discoveries addresses a wide range of theoretical concerns and presents a wealth of empirical examples from the history of science. Brannigan has something to say about everything: epistemology, relativism, error and deception, priority disputes, multiple discoveries, the supposed neglect of Mendelian genetics, the impact of the telescope on astronomy, and even the discovery of America. Herein lie both the strength and the weakness of the book; the author deals with many issues, but not equally well with all of them. I find the case for the attributional model of scientific discovery completely successful, but some of the secondary arguments are less persuasive. For example, Brannigan asserts that erroneous models of scientific discovery have prevailed because they reconcile the observation that the human agency, the "genius" of individual scientists, is necessary to scientific discovery with the objectivist epistemological premise that discoveries arise from the natural world. On this point, Brannigan's evidence is weak and his prose murky. Nevertheless, even when he confuses or fails to persuade, Brannigan is stimulating. There is so much in this book that no one who is interested in the sociology of science or the sociology of knowledge will want to miss it.

Negara: The Theatre State in Nineteenth-Century Bali. By Clifford Geertz. Princeton, N.J.: Princeton University Press, 1981. Pp. xii + 295. $18.50 (cloth); $5.95 (paper).

Peter K. Manning
Centre for Socio-Legal Studies and Balliol College, Oxford, and *Michigan State University*

This allusive book incorporates a subtle nudging of political theory, an essay in dramaturgical analysis, a historical ethnography, and a poetic vision. It claims much about the nature of poetics, politics, and analysis; and it compresses, cites, and displays an almost inundating quantity of daunting material: references, explicanda, asides, comments, and esoteric multileveled arguments carried on in footnotes that would make Borges envious (see, e.g., p. 245, n.). The section labeled "Notes" occupies

fully 159 pages, and, with bibliography and index, complements 136 pages of exposition.

It is not easy for common scholars like myself to evaluate this kind of work using the facts that are presented. What could be written that we could see as untrue? *Negara: The Theatre State in Nineteenth-Century Bali* could be reviewed as a narrowly argued contribution to the social anthropological analysis of Balinese culture. On these grounds, few would be qualified to comment on it. The book, however, claims to be much more, and the claims, at least, are quite rewarding to consider.

Clifford Geertz sets out, on the basis of his fieldwork and a distillation of the literature, to ". . . construct . . . a circumstantial picture of state organization in nineteenth-century Bali" (p. 7) and to draw from that some general guidelines for ordering "pre- and protohistorical material in Indonesia (and beyond it, Indic Southeast Asia) generally" (p. 7). He avers to reshape political theory, rejecting the notion that bestial instincts are merely contained by institutions or by normative devices. This Hobbesian metaphor, according to Geertz, is flawed: ". . . in Bali [this idea's] absurdity is patent. The passions are as cultural as the devices; and the turn of mind—hierarchical, sensory, symbolistic and theatrical—that informs the one informs the other" (p. 124). Court life, with the king as an iconic representation of that, and "the ritual life of the court, and in fact the life of the court generally, is . . . paradigmatic, not merely reflective of social order" (p. 13). One does not contain the other, they are two sides of the same coin.

Geertz stylishly repeats throughout the essay that ritual displays the state. Drama is all that *is* at the level of the state. Conversely, the delicately administered irrigation systems, surrounded by small "villages" which are laterally organized and vertically organized by complex, ritually defined kinship, the focus of the administrative apparatus, are said to be in conflict with the symbolized center. "Balinese politics can be seen as stretched between two opposing forces: the centripetal one of the exemplary state ritual and the centrifugal one of state structure" (p. 18). Ritual acted out cultural unity, while the material arrangements produced segmental, quasi-independent micropolitics. Describing, on the one hand, the complex interdigitation of local and regional authority—of ritual, public life, and irrigation administration—while capturing the tension among cultural, social, material, and ritual domains, kinship and authority, and, on the other, etching conflicts between temple ceremonies, local authority, trade and the state—is a demanding textual production. Perhaps the pellucid prose obscures, as, for example, when Geertz writes: "the extravagance of state rituals was not just the measure of the king's divinity . . . it was also the measure of the realm's well-being. More important, it was a demonstration that they were the same thing" (p. 129). "The king owned the country as he ruled it—mimetically; composing and constructing the very thing he imitated" (p. 128). Elegance of expression, often attributed to Geertz, is often apparent in excess.

In this brief, densely argued essay, Geertz has taken on theories of Indic civilization and religion, and ecological power theories of the state (Marx, Wittfogel, Lenin) while carrying structuralist clothing in Weberian luggage. Parenthetically, it is surprising that in the vast notes section (120 pp.) attention is given neither to such precursors as Burke, Shils, and Goffman, nor to A. M. Hocart and Louis Dumont. The theory that a ceremonial or exemplary center of society exists, the purpose of which is to image all that should be imagined, was first advanced by Hocart and elaborated in an analogous fashion in India by Dumont. On the one hand, the claims, aims, and scope are grand. On the other, specific assertions are indisputable yet arguable in nature. For example, consider his assertions concerning the role of the *Karrula*. According to Geertz, it invoked two analogous cultural equivalents of ritual service and military support: "He was not a tenant, a serf, a servant, or a slave. He is not even what *faute de mieux,* I have been calling him, a subject. He was a stagehand, spear carrier, and claqueur in an endless political opera" (p. 65). Would this claim be viable in an alternative interpretation of the material basis of this culture?

Geertz weaves together historical fact, current data, and observation based on his own extensive fieldwork in the area and arranges them to represent the substance of the precolonial state as drama, myth, and ceremony. In so doing, he illustrates the isomorphic aspects of "practical" instrumental politics and expressive actions. Apparently the ritual character of actions, and the state as an organized spectator, cannot be isolated here from the exercise of power. The paradox, of course, is that Balinese conceptions of power are apparently based on an admiration of the display of disciplined, controlled passivity. Thus, the state is not an instrumentality requiring ceremonial marking because social and cultural symbolization processes constitute the state.

Would such an argument work in a society in which religion and politics serve to *symbolize action* and attempt to freeze class hierarchies? Infrastructure and ideology may not cohere. Further, could it not be claimed that the disintegration of the irrigation groups of the local villages was a direct result of the power of the central families, and that center control meant, as it did in medieval feudalism, competition among peripheral, potential successors?

This essay, sprinkled with words from several languages, exemplifies a comparativist approach to social structure, emulating Weber's China/ India comparison and challenging the steamroller Marxism of world-system "theory" as well as much conventional social science wisdom about the nature of social structure, politics, and the material world. The power of the book lies in its implicit criticism of the culture-bound Anglo-American vision of politics still dangling and swinging from the works of Hobbes, and its evocation and, indeed, celebration of the polity as expression and expression of power. Can politics be other than the dancing of dreams?

Wilhelm Dilthey: The Critique of Historical Reason. By Michael Ermarth. Chicago: University of Chicago Press, 1978. Pp. xiv + 414. $19.00 (cloth); $9.50 (paper).

Jeffrey Barash
University of Chicago

In his book *Wilhelm Dilthey: The Critique of Historical Reason,* Michael Ermarth asserts that Dilthey's philosophy is not only of interest to the historian of late-19th and early-20th-century ideas, but that it has general relevance for present-day intellectual concerns. In our age of information theory, of advanced computer techniques, and of highly developed quantitative methods of analysis, Dilthey's theories, for Ermarth, are still of cardinal importance for the human sciences. "At the very least," Ermarth writes, "it can be said that no serious effort to advance the theoretical discussion of the human sciences can afford to ignore Dilthey. It is the purpose of this study to establish this point by means of a comprehensive exposition of his thought" (p. 4). Ermarth's perceptive investigation of Dilthey's theories, which draws on a wealth of unpublished texts from the Dilthey literary estate, fulfills this aim amply.

In treating Dilthey's thought Ermarth is sensitive to the many-sidedness of this philosopher's endeavor. Dilthey wrote in the fields of psychology, history of philosophy, epistemology, biography, and literature, to name only the predominant areas of his interest. Ermarth does not deal exhaustively with Dilthey's writings in each of these domains but investigates the philosopher's diverse concerns chiefly in relation to his overarching theory of the human sciences, or "Geisteswissenschaften." According to Dilthey's own definition, such a theory comprehends the "totality of sciences that have social-historical reality as their object" (W. Dilthey, *Einleitung in die Geisteswissenschaften,* vol. 1 of *Gesammelte Schriften* [Stuttgart: Teubner, 1971], p. 4; my translation).

In arguing for the contemporary relevance of Dilthey's thought in this area, Ermarth enters into a discussion that has long engaged Dilthey scholars: whether a theory of the human sciences is coherently elaborated in Dilthey's philosophy, in spite of the range of his interests and his lack of an *esprit de système.* Although Ermarth admits that Dilthey's ideas are not free of contradiction and that there is "some truth in the judgment that his merit lies more in what he began than in what he finished" (p. 349), he argues that Dilthey did provide a coherent vision of the *Geisteswissenschaften.*

The picture of Dilthey that emerges in his work is that of a thinker who extended the scope of his ideas and continually revised them in an attempt to do justice to the limitless diversity and richness of a uniquely human social and historical reality. As Ermarth points out, Dilthey strove constantly to approach this reality through a "foundational science" (*Grundwissenschaft*) by which the heterogeneous data of human social

483

and historical experience could be interpreted faithfully and made fruitful for the various disciplines of human study.

To be sure, Dilthey's vision of a foundational science for the *Geistes-wissenschaften* changed over time. Ermarth's book performs a special service in placing Dilthey's idea of a foundational science in intellectual context and in showing, within this context, how his idea developed. Ermarth points out that Dilthey first focused on the psychology of immediate lived experience (*Erlebnis*) as the basic science of human reality. Later, he shifted emphasis and sought the ground for such a science more directly in the patterns of mediation and objectification of lived experience within a cultural framework. Here a consideration of the linguistic and institutional cadre of experience (termed "life-experience" or *Lebenser-fahrung*) became of more central concern than in the earlier stages of Dilthey's career.

Ermarth's book is especially strong in its analysis of Dilthey's mature concept of understanding (*verstehen*), which was the fruit of his lifelong preoccupation with the theoretical foundation of the human sciences. The concept of understanding incorporated Dilthey's reflections concerning the interpretation of language or of other modes of patterning experience (gestures, social conventions, etc.) that take form in the field of human culture. Through Ermarth's insightful investigations we see that this concept has lost nothing of its relevance; a number of topics Dilthey analyzed in this regard are still at the heart of contemporary debate in hermeneutics; among these the following are of special importance: (1) Dilthey's theory of the relationship between understanding a text and understanding the historical context in which it emerged; (2) Dilthey's assumption that understanding of a text is not limited to reconstruction of an author's intentions in writing it; and (3) Dilthey's distinction between understanding in the human sciences and explanation (*Erklärung*) in the natural sciences.

Ermarth is generally favorable toward Dilthey's ideas—at times, too favorable. Although he does offer some critical remarks, especially in the conclusion of his work, he nearly always defends Dilthey against his opponents. A more impartial consideration of the opposition that Dilthey's theories have aroused, especially in the half-century following his death, would have helped provide a more balanced assessment of Dilthey's legacy (e.g., pp. 356–57).

In spite of this shortcoming, however, one cannot put down Ermarth's book without sensing that he has made a major contribution to Dilthey scholarship. His work provides deep insight into Dilthey's "notoriously enigmatic" thought (p. 4) and brings that thought to life for present-day concerns.

Spectacles and Predicaments: Essays in Social Theory. By Ernest Gellner. Cambridge: Cambridge University Press, 1979. Pp. viii + 385. $34.00.

John O'Neill
York University, Toronto

This is the fourth volume of Ernest Gellner's essays. It was put together by I. C. Jarvie and J. Agassi, who do not believe in everything Gellner writes but who do, presumably, believe that most of it has a permanent claim on our attention. I am not so sure of this, but I would agree with them that Gellner is thought-provoking and witty enough to be a treat to read. He is also infuriating, misleading, and quite often misses the mark. Well, since there are few if any sociological angels, we must take Ernest as he is—an important, if impertinent, Popperian.

Since much of Gellner's good derives from the reader's enjoyment of his lively way of presenting large issues and revered thinkers, I can only urge the reader on. In fairness to less exciting writers, it should be remarked that Gellner has a penchant for controversy and can count on a variety of outlets for the exposition of his views. Gellner does not write for promotion and tenure, thank God! In North America, we are likely to be puzzled by the variety of Gellner's interests and by the array of journals, magazines, and newspapers in which he is able to carry on a high level of academic discourse in which philosophical, sociological, and political themes are elegantly interwoven.

Spectacles and Predicaments contains 19 essays organized around three themes: how we understand ourselves, how we understand the nature of scientific truth, and how we understand the polity in which our understanding of the first two issues is most likely to be fostered. On the first theme, chapter 3, "A Wittgensteinian Philosophy of (or against) the Social Sciences," and chapter 6, "Notes towards a Theory of Ideology," are well-developed arguments with an enduring place in their respective literatures. On the second theme, Gellner writes at considerable length on pragmatism with special reference to Quine, in whom he discovers a metaphysics and politics so far unnoticed by U.S. observers. Chapter 11, "The Last Pragmatist, or the Behaviourist Platonist," is a tour de force in grappling with Quine. Its plot is perhaps revealed in the title of chapter 12, "Pragmatism and the Importance of Being Earnest" (Gellner?). In my view, then, these are the four substantial essays in parts 1 and 2. The essays in part 3 deal with nationalism, the laissez-faire theory of the state, some problems of the British welfare state, the Czechoslovakian state, Nepal, and an argument with Raymond Aron over the relation between revolution and liberalization. There are interesting remarks in all of these essays. But I think the more enduring things—or should I say "the lasting pleasures"?—are to be found in the first two parts of the collection. A good way to get at the menu is to use the name and subject index to see what

Gellner has in store for you. He is a bit short on French and German imports, but then Giddens runs the store for them.

Of course, Gellner is not all style. The title, *Spectacles and Predicaments,* is meant (I believe) to convey an opposition between two philosophical camps. One cannot say it reflects two opposing *sociologies,* since one of the philosophical camps considers society is best understood philosophically and without the aid of sociology. The latter view is generated by post-Wittgensteinian, ethnomethodological, and pragmatist idealism, as well as by Marxist historicism and any form of cultural relativism from Winch to Feyerabend. Gellner, it may surprise you, is *for* sociology—though not in the same way C. Wright Mills, Gouldner, or Max Weber is. He rejects idealism wherever it is to be found (he is a considerable sleuth of such sloth) because it treats society as a comfortable spectacle, the sort of thing we have in our blood if we have the right blood. He espouses sociology because, unlike the idealists, he is a man of predicaments who has seen across the "Big Ditch" between nonindustrial and industrial societies. The benefits of being on our side of the ditch—assuming we keep a careful watch on our internal weakness for idealism, relativism, existentialism, and Marxism—is that we have a conception of scientific truth and of sociological problems which, when it is combined with an aversion to statism, puts a healthy distance between us and the tribal past (pp. 336–37, 339–40). The tribal past may as well be accorded a charitable relativism when it comes to collecting its samples and curios, but there can be no question of our taking a second place either from romance or from guilt. Any such temptation would merely plunge us into the dark night of barbarism, deprived of the science, technology, and medicine that are all that stand between us and universal chaos.

Gellner, then, is an unrepentant Popperian, still fighting off the demons of the Open Society. Having abrogated to Popperianism all there is of critical sociology, Gellner's arguments leave out much that other sociologists would consider relevant. The only remedy I can think of is to try to buttonhole Gellner in an argument. But be sure you are as quick as he is. Alternatively, Jarvie and Agassi should consider an edition in which some of Gellner's whipping boys are allowed to speak. Many of them are just as smart as he is. The readers would then get a better run for their money.

A History of Anthropological Thought. By Edward Evans-Pritchard. Edited by André Singer. Introduction by Ernest Gellner. New York: Basic Books, 1981. Pp. xxxvi + 218. $15.00.

James Clifford
University of California, Santa Cruz

At his death in 1973, Edward Evans-Pritchard was revising for publication a series of lectures on the history of his discipline, lectures he had

given for many years at Oxford. In *A History of Anthropological Thought,*
André Singer has now ingeniously stitched together five completed seg-
ments, notes, prefaces, obituaries, and other published passages to make
a book of 15 chapters, each devoted to an author. An appendix supplies
sketches for 10 more.

Evans-Pritchard is less a historian than a genealogist. Showing only
passing interest in historicist approaches, he presents and refutes his
chosen thinkers as if they were contemporaries. From Montesquieu, an
intellectual filiation passes by way of Comte and Durkheim (with parallels
in the Scottish Enlightenment) to the social anthropology in which Evans-
Pritchard is, himself, ambivalently located. On Radcliffe-Brown and Mal-
inowski, his most problematic precursors, we have scraps of what appear
to be highly critical assessments. The final chapter, on Robert Hertz,
provides a few systematic paragraphs on the "comparative method" Ev-
ans-Pritchard espouses, obliquely, throughout the book. Readers will do
well to begin this meandering history at the end.

In his discussion of Hertz, Evans-Pritchard identifies a "fundamental
aim which has remained the same from Montesquieu to Durkheim: to
seek the general in the particular by comparing different societies and to
try to understand the general in terms of the interrelationship of insti-
tutions and of certain uniform tendencies in their historical development"
(p. 170). From the universalist, deductive schemes of the 18th century,
Evans-Pritchard traces, via J. F. McLennan and William Robertson Smith,
a progressive "limitation of the field of comparison." Hertz's essays on
mortuary rituals and the symbolism of Left and Right represent "a cul-
mination of two centuries of sociological thought in France," a tradition
with which Evans-Pritchard cautiously aligns himself. Hertz is made to
embody "the movement towards a truly scientific study of social phe-
nomena . . ." (p. 172).

"Truly scientific" comparison presupposes the attitude, if not the actual
experience, of a functionalist fieldworker; it densely interconnects and
translates local symbols and actions. But the comparatist must sift the
general from the particular, "the social fact from its cultural form" (p.
173). Evans-Pritchard shows little interest in just how this sifting is ac-
complished. He is less concerned with ethnography than with a compar-
ative sociology that builds on abstractions from separate societies in an
attempt to elicit cross-cultural generalizations (not explanations). This
middle-level theory is Evans-Pritchard's *telos,* the "comparative method"
toward which most of the figures in his history are striving and against
which most are found wanting.

The resulting genealogy has a limited scope, and it would be fruitless
to belabor its omissions. Most obvious is the almost complete absence of
Germans and Americans in this odd gathering of Scots and French. This
skewed "history of anthropological thought," is, perhaps, best seen as a
personal document, a "troubled" scholar's "dialogue with his predeces-
sors," as Gellner puts it in his introduction. But the book is a patchwork
of fragments composed over a period of a quarter of a century; the seams

are often invisible, making it difficult to tell what piece was written when. Thus it is a frustrating scource for the development of Evans-Pritchard's ideas. Singer, generally respecting chronology, ends the book with Hertz (who stands, explicitly, for the entire Durkheimian tradition), being offered as the direction for sound future work. But in the immediately preceding chapter, Evans-Pritchard firmly reads Durkheim out of the ranks of science. Our confusion is compounded by the fact that the final chapter was composed at least 25 years *before* the preceding section. To make things worse, the pages of overview in the Hertz discussion seem to have been added later. Nowhere is this sort of editing explained. The form given to the overall narrative by the final chapter is inexplicably undermined by the adjoining Durkheim criticism. Must we assume that, given time, the author would have changed the conclusion? The book's tenuous coherence begins to dissolve.

Should this collection have been published at all? For those interested in Evans-Pritchard's development, it muddies the waters. As a serious study it is a far from adequate exercise in intellectual history. Its presentist and tendentious discussions vary widely in tone and depth. Some parts are little more than annotated reading notes (and this is true even of the chapters Evans-Pritchard had apparently completed). Readers interested in a distinguished anthropologist's thoughts on the background of crucial issues in his field will find little in these pages that is not said with more rigor and panache elsewhere in his work.

But the book exists, testimony to a laudable if questionable piety, a further document from a long and increasingly eccentric career. Evans-Pritchard never seems quite to believe in the "comparative method" he espouses. Having renounced Radcliffe-Brown's natural scientific paradigm in the late 1940s, he seems to wander in the borderlands—between skepticism and belief, between the dream of a rigorous comparative sociology and the claims of rhetoric. He looks to history, and these are his gleanings. At his best he is a trenchant critic, interestingly perverse. He demolishes reputations (Durkheim, Frazer); he rereads despised classics generously (Comte, Lévy-Bruhl); he brings forgotten names to mind (Lord Kames, John Millàr, Condorcet, Van Gennep, Franz Steiner, A. M. Hocart). At his worst he is superficial, haphazard, dismissive. On balance, one would rather remember Evans-Pritchard for his better-honed historical-methodological polemics of the 1950s and early 1960s and for his ethnography.

Aging: Social Change. Edited by Sara B. Kiesler, James N. Morgan, and Valerie Kincade Oppenheimer. New York: Academic Press, 1981. Pp. xxiv + 631. $24.50 (paper).

Jeylan T. Mortimer
University of Minnesota

The 22 chapters in *Aging: Social Change,* edited by Sara B. Kiesler, James N. Morgan, and Valerie Kincade Oppenheimer, were presented at a workshop, "The Elderly of the Future" (May 3–5, 1979), organized by the National Research Council's Committee on Aging and sponsored by the National Institute on Aging. Among its 30 contributors are sociologists, psychologists, political scientists, demographers, and anthropologists. But in spite of its great variety of perspectives and methodologies (simulation, econometric analysis, demographic projection, survey research, anthropological field observation), the work has some major identifiable integrative themes.

Of foremost importance is the pervasive emphasis on the need to link macro and micro phenomena in studying aging and the effects of the aged on society. For example, the well-being of the elderly in the future will depend on economic trends, government programs, legal change, and corporate policy, as well as on attributes of the elderly themselves—their education, attitudes, values, information, decision-making processes, and family ties. James Morgan calls for the collaboration of behavioral scientists and econometricians to strengthen models of economic and behavioral functioning.

Changing demographic trends suggest major differences between cohorts of elderly people and their social environments. According to E. A. Hammel, K. W. Wachter, and C. K. McDaniel's demographic projections, the elderly of the near future, because of their previous relatively high fertility (especially during the baby boom years), will have substantial resources for social and material support. These researchers' simulation of population characteristics to the year 2000 highlights the competition between the elderly and their grandchildren for the resources and attention of the middle generation. According to S. Kiesler, the growing proportions of elderly people in the population will foster positive attitude change. Using motivational and cognitive theories, she argues that, as a result of increasing contact with older people, significant change in behaviors will occur, generating self-justifying attitude change and the elaboration of stereotypes.

Several authors point to the increasingly important role of government in determining the constraints and opportunities facing the aged and foresee the elderly as a potent political force in the future. R. G. Fox describes advanced industrial societies as welfare states that are highly intrusive in regulating the private affairs of their citizenry. The welfare state legitimates appeals for higher living standards and equity on as-

criptive grounds, creating a situation ripe for the mobilization of the elderly as a political constituency. W. R. Scott develops this theme in his description of rationalized reform processes in modern societies: a disadvantaged category is identified; forces calling for reform are mobilized; inadequacies of the knowledge base upon which reform rests are recognized; and clients' needs are seen as increasingly independent of demands for services. But reform organizations may be more concerned with following the directives of institutional sponsors (e.g., area agencies on aging and nursing homes) than with enhancing the quality of actual service provision.

Growing political strength of the elderly is predicted, given the decline in political partisanship and changing demographic patterns which extend the base of support. Moreover, in comparison with previous cohorts, the aged are better educated, wealthier, and more "age conscious" (Cutler). These trends could dramatically increase political opportunities for them in future years (J. A. Schlesinger and M. Schlesinger). Some argue that gerontologists will (or should) play an active role in this political awakening (W. R. Scott). To Fox, gerontologists' emphasis on the victimization of the aged contributes to the necessary ideological grounding of political movements of the elderly. The problems of the old are seen as resulting from discrimination, prejudice, forced retirement, and unresponsive institutions, rather than from the process of aging or the deficiencies of elderly people (J. Cooper and G. R. Goethals).

At a more microanalytic level, the elderly are shown to be crucially affected by their own past experiences. D. Sears, in his review of research on political attitudes through the life cycle, concludes that attitudes are highly susceptible to external influence in the early years but have considerable stability later. Consistent with this view, J. Pfeffer argues for persistent cohort effects in formal organizations. Cohorts are "imprinted" with the ideas, values, and skills obtained early in their careers. This results in conflict among cohorts and changes in leadership style as new cohorts come into power. But in spite of the liberalizing effects of cohorts on attitudes over time, Davis demonstrates the importance of a recent conservatizing period influence.

Differentiation of life-styles among the currently middle-aged leads to the prediction of greater heterogeneity among the elderly in future generations. First, there may be greater economic inequality (J. Treas). While some depend entirely on social security, others have one, or even two, private pensions in addition (as do some dual-earner couples). S. Cobb and J. Fulton predict greater differentiation of the elderly population with respect to health because of earlier life-style variation by social class. Recognizing that earlier attitudes and experiences largely determine the attributes of persons later in life highlights the importance of studying the patterns set in earlier life stages. This holds true whether one is studying health and socioeconomic outcomes, political attitudes, life satisfactions, or other intellectual, affective, and behavioral attributes (R. Sears).

Social support is given predominant emphasis as a contemporaneous source of satisfaction among the elderly. As the individual moves through life stages and their associated roles, a convoy of social support is carried along (R. L. Kahn and T. C. Antonucci). Researchers should specify the features of convoys in different life phases to identify what is normal and beneficial and to target "at-risk" populations. C. Dunkel-Schetter and C. B. Wortman consider parallels between victimization and aging (which involves losses like retirement, widowhood, physical decline, and illness). The elderly need support for "ventilation" (the airing of problems, which helps in their resolution) and "validation" (reassurance that their reactions to problems are normal). But these authors point out that the situation of the elderly often arouses reactions in others that prevent them from giving the kinds of support that are needed. Feelings of vulnerability to the same fate and helplessness may engender avoidance and the failure to communicate openly.

Several authors conclude that support from age peers is more efficacious in promoting the elderly's adjustment than support from younger kin. Dunkel-Schetter and Wortman stress that support is needed from people with similar problems—who will find the elderly's "ventilation" interesting and important, and who can provide validating information. J. Keith describes the distinctive attitudes about death in nursing homes and reiterates the importance of supportive and understanding age peers. Keith's anthropological perspective points to similarities in these age-homogeneous communities across societies. They establish their own rules, status distinctions, and rites of passage that counter the rolelessness, stigmatization, and dependency of the elderly in the wider society.

It is generally agreed that greater attention should be given to age, particularly in the fields that have traditionally neglected it. As a case in point, Pfeffer shows the effects of tenure distributions on formal organizations: on control and governance processes, leadership succession, power, the size of the administrative component, and problems of the "career plateau." According to his analysis, increasing predominance of long-tenured workers decreases the infusion of new ideas and reduces organizational capacity to change. S. Stewman's simulation shows the increasing prevalence of this situation given current managerial practices and worker exit behavior.

Along with suggested research agendas, intervention strategies are proposed—to strengthen the elderly's support networks (Dunkel-Schetter and Wortman, Kahn and Antonucci), to increase the capacity to live independently (B. J. Soldo, R. J. Struyk), and to modify organizational roles to enable continued commitment and contribution (Pfeffer). J. Treas points to contradictions in the social security system that reformers will confront in the future.

The chapters are generally well written, interesting, and informative. Researchers will value this book for the cogent summaries of current theoretical and empirical developments and the identification of problems requiring further investigation. Policymakers and practitioners will profit

from the discussions of public issues and programs. This excellent overview of current thought on aging in the social sciences is appropriate for use in graduate seminars and upper-level undergraduate classes on aging, gerontology, and life-span development. The authors, editors, and reviewers, the Committee on Aging, and the NIA sponsors deserve congratulations for an important contribution to the social science literature on aging.

Transitions and Social Change: The Early Lives of American Men. By Dennis P. Hogan. New York: Academic Press, 1981. Pp. xiv + 232. $26.50.

W. Parker Frisbie
University of Texas, Austin

Transitions and Social Change, a generally useful addition to the Academic Press Studies in Population series of research monographs, focuses on three "transition events": completion of schooling, attainment of a full-time civilian job, and first marriage, as they pertain to American men born from 1907 to 1952. Data are drawn from the Occupational Changes in a Generation II survey. Although these three events, either singly or in combination, have been the subject of much prior research, Dennis P. Hogan's analysis differs from most earlier work in adopting a life-course approach. (In this the author is heavily and explicitly indebted to the work of Glen Elder.)

In general, life-course research combines an interest in cohort and period effects. Hogan defines the relationship between the two in this fashion: "Thus, the historical milieu experienced by an individual is determined by birth cohort, but the nature of the experience is filtered by the individual's position in the social structure" (p. 6). The author's description and analysis of the transition from childhood to adulthood are more successful in the descriptive than in the analytical portions.

Description of the above transition is accomplished by tabulating cohort differences in age at first quartile, median, and third quartile for completion of schooling, first job, first marriage. Where the sequencing of events is in that specific order, the transition is seen as normative. All other permutations are considered to be "nonnormative," with marriage prior to completion of education characterized as "extreme nonnormative." Perhaps the major trend observed was a shortening of the transition interval (somewhat arbitrarily measured by the interquartile range) as one shifts attention from older to younger birth cohorts. A likely explanation of this decline in duration is provided by a lengthening period of school enrollment coupled with a declining marriage age for 75% of a cohort. Tabulations are also presented which show variation in sequence and length of transition by a large number of structural variables combined into six factors representing modernity (strong loadings on educa-

tional attainment, as well as on economic and urban variables), social welfare expenditures, unemployment, military service demands, economic growth, and "demographic transition" (mortality and fertility measures).

Following the description of patterns and trends is an analysis of individuals, as opposed to cohorts, which applies log-linear techniques to discover whether parental occupational status, size of community of origin, and ethnic/racial ancestry can explain differences in the timing and ordering of the transition to adulthood. Finally, presumed consequences of timing and ordering for occupational and earnings attainment are treated at various career stages, that is to say for those men who began a first job 0–4, 10–14, and 25–29 years prior to the survey.

Problems with the analysis emerge at several points. There are 15 large matrices involving a transition parameter cross-classified by cohort, educational level, military service, and social background variables that serve as inputs for the log-linear analysis. The result is often extremely short on interpretation in comparison with the number of tabular materials. For example, in dealing with the relationship of size of community of origin to the transition parameters, about two pages of text are used to describe and interpret eight tables. Whether all of the tables are really necessary is the author's decision, and it is probable that the imbalance is caused by space limitations. Nevertheless, the treatment is far too superficial to give much insight into how the voluminous set of results fits (or does not fit) the conceptual framework guiding the analysis.

Another difficulty is particularly apparent in the sections dealing with the consequences of different transition patterns. The author attempts to investigate the *process* of occupational and earnings attainment by comparing, for example, men who have been working for 10–14 years with those who have been working for 25–29—this in lieu of life-history data. An illustration of the type of conclusions Hogan draws is his statement that initial effects related to time of the first job transition "persist through the first 14 years of the career," but diminish to insignificance by the 25–29-year mark (p. 191). The obvious problem with such an interpretation is that there is no way of knowing whether the initial advantages/disadvantages of men at the 10–14-year career mark will maintain themselves, decline, or increase over the next 10–15 years. Nor is it known whether those men who have a 25–29-year career were in the same position at the 0–4- or 10–14-year stage as those men currently in these early stages. To his credit, Hogan is quick to acknowledge the tenuous nature of his inferences about intercohort differences, but the recognition of a problem is not the same thing as a solution.

In summary, it is only fair to point out that a frequent source of weakness in this volume is the extremely broad scope of the analysis—the very aspect that accounts for much of its strength. Hogan not only carries on the investigation at two levels (cohort and individual) but also attempts to discern both causes and consequences of transition events while placing the entire set of relationships in proper historical context. I am inclined

to agree with the author's own assessment that, although the analysis succeeds only occasionally, it will be appreciated by professionals and students across a wide range of disciplinary interests.

The Population of the South: Structure and Change in Social Demographic Context. Edited by Dudley Poston, Jr., and Robert H. Weller. Austin: University of Texas Press, 1981. Pp. ix + 307. $25.00.

Neil Fligstein
University of Arizona

The study of the demography of the South has a long history. Because the region has differed traditionally from the rest of the country in a number of dimensions, much sociology on these issues has been produced. *The Population of the South,* edited by Dudley Poston, Jr., and Robert Weller, a set of papers written for the Southern Regional Demographic Group's meeting in 1976, continues in this tradition. The papers try collectively to address the issue of whether the South still forms a distinctive region in the United States and conclude generally that regional convergence has occurred. On the whole, the papers are informative, although none is particularly innovative either theoretically or methodologically. The authors' relatively narrow demographic focus describes much and explains little.

Poston, W. Serow, and Weller begin the volume with a general overview of demographic change in the South since 1870 and a summary of the other papers. They note the revival of interest in the South in conjunction with recent attention paid to the rise of the sun belt. Their conclusion is that while the South is increasing its share of population, at the same time, Southern fertility and mortality rates are converging with those of the North. They document the well-known turnaround migration to the South, noting that much of that migration is to Texas and Florida. The next three chapters present more detail on these basic patterns. R. Rindfuss shows that convergence in fertility rates occurred during the baby boom and resulted from the fact that Southern women (in particular, older and less educated women) increased their fertility less than non-Southern ones. H. Rosenberg and D. Burnham present extremely detailed mortality statistics showing impressive convergence of mortality trends. The South closed this gap for three reasons: better medical care, better living conditions, and a decline in the number of blacks. D. Sly's paper on migration describes the well-known patterns of migration from the South and within the South.

Poston attempts to provide an explanation of Southern population growth from 1970 to 1975 utilizing an ecological explanation. Unfortunately, this perspective is not enlightening as an explanation of migration to Southern counties. The models presented are clearly misspecified; they contain no

494

change variables as independent variables, no lag migration variables, and no measures that capture quality of life, urban growth, or any other factors thought to explain the phenomenon. O. Galle and R. Stern document the stability of the Southern urban hierarchy from 1950 to 1970 and show the dramatic rise of Houston to a city of national importance. J. Singelmann describes the rapid decline of Southern agriculture and the increase in manufacturing and service industries from 1920 to 1970. The industrial and occupational structure of the South has come to resemble the non-South in a remarkably short period of time. Serow tries to use economic arguments to explain why the South was transformed and how this affected demographic patterns. He argues that low taxes and low wages were preconditions of Southern development, and the highway system was the catalyst for explosive growth. Further, Serow suggests this process has now been set in motion and will continue.

S. Clarke tries to establish how political decisions have and could affect population change in the South. Her discussion is wide-ranging and quite interesting. G. Myers presents a summary paper that concludes that the South has been integrated demographically into the nation and that Southern regionalism, while once a useful concept, is now less viable because of regional convergence. He suggests a research agenda that includes studying urbanization, the elderly, issues of race, and the relations between the South and Central and South America.

The strong point of this volume is its massive summary of data on trends in Southern demography. Its major weaknesses are threefold. (1) Much of what is said here has been said before. (2) The data analysis is unsophisticated and generally does not present causal modeling of any process. (3) This volume is almost totally descriptive and never addresses the issue of how and why Southern agriculture was transformed from a tenant system to a machine-intensive system. Further, the authors (with the partial exception of Serow and Clarke), never really explain the rise of the sun belt and never present any causal modeling of that process. That the South has, in many ways, converged with the rest of the country is something we all have known. Why that is so is a book that remains to be written.

Theories of Trade Unionism: A Sociology of Industrial Relations. By Michael Poole. London: Routledge & Kegan Paul, 1981. Pp. xi + 265. $35.00.

Daniel Cornfield
Vanderbilt University

In his *Theories of Trade Unionism,* British sociologist Michael Poole presents a comprehensive synthesis and intellectual history of the disparate schools of theoretical and empirical research on trade unionism.

The 20-page bibliography of mostly British and American research (over half of the citations are references to British research) attests to the breadth of the work. Noting that "there have been remarkably few attempts to draw together the main strands of classical and modern approaches to labour theory and to integrate these conceptions into" sociological propositions (pp. 1–2), the author, a senior lecturer at the University of Wales, seeks to fill this gap.

The chapters proceed chronologically, showing how distinct, classical theories of trade unionism have become increasingly interwoven into modern theories of trade unionism. Five classical schools of thought—moral/ethical, revolutionary, psychological, economic, and political—are presented in the first chapter, and chapters 2 through 4 explain how John Dunlop's structural-functional theory and the Oxford, liberal-pluralist school synthetically incorporate the earlier theories. While impressed by the complexity of these modern theories, Poole notes their weaknesses. Summarizing the extant debate on structural-functionalism, he criticizes the approach for, among other faults, its "indiscriminate multi-causality" (p. 33) and its presentation of untestable hypotheses. The Oxford school is criticized for its narrow focus on the effect of bargaining structures on trade union behavior as opposed to seeing these structures as intervening variables between societal influences and unionism.

These inadequacies lead Poole to develop an analytical framework in chapters 5 through 8 in order to promote further research on trade union action and behavior. Power is a central explanatory variable in the framework. After reviewing Weberian, Marxian, and contemporary theories of power and unionism in chapter 5, he emphasizes the importance of distinguishing between power resources and factors affecting the exercise of power in order to explain union behavior. In concluding his chapter 5 discussion, Poole claims that, despite its centrality in a theory of unionism, power, stripped of the structural and cultural determinants of its distribution and exercise, provides a limited explanation of trade union behavior.

To place power in a broader context, Poole draws on recent research in considering a variety of independent variables that affect the exercise of power by unions and, therefore, trade union behavior. In chapter 6, he discusses the effects of such primary structural constraints on power exercise as politicoeconomic structure, technology, and division of labor, as well as secondary constraints, including class and occupational structure, enterprise size, community structure, and microstructural elements. The union is treated as an organization nested in the context of these structural constraints in chapter 7, where he examines the effects on power exercise of such organizational variables as administrative structure, leadership, and members' expectations. Using Weber's four types of social action in chapter 8, Poole examines how such subjective factors as imagery and orientation affect the goals and objectives of trade unions.

By organizing the discussion according to independent variables, and by adopting a general and vague dependent variable, "trade union action

and behaviour" (p. 142), Poole obscures the purpose of his analytical framework. Some 15 dependent variables which encompass such divergent topics as union density, strike propensity, shop steward behavior, and participation in union affairs are mentioned and presumably serve as dimensions of trade union action and behavior.

At the very least, trade union action is directed toward two fundamental tasks: recruiting new union members and improving the socioeconomic livelihoods of union members. A theory of trade unionism sought to provide a separate explanation for each of these facets of trade union action and behavior, as each is affected by different causes. For example, the causes of the post–World War II decline in union density in the United States, neglected by Poole, might include the decline in blue-collar jobs and status anxiety among white-collar employees who refuse to join unions because they perceive them as low-status organizations. By contrast, the nascent forms of Western European industrial democracy in the United States, as seen in the union representation on the corporate boards of Chrysler and Pan American World Airways, may derive in part from other causes, such as adverse macroeconomic conditions, cross-cultural diffusion of ideas, and concern over profitability and job security. Fusing these causes in one analytical framework as causes of trade union action and behavior reduces the explanatory power of a framework which should distinguish more clearly between different aspects of trade union action and behavior.

Theories of Trade Unionism accomplishes a comprehensive review of the labor literature, although it does so partly at the expense of the intent of the author's synthetic, analytical framework. The work is valuable for sensitizing scholars to the complex causes of the past patterns and the destiny of the labor movement.

Bargaining, Power, Tactics, and Outcomes. By Samuel B. Bachrach and Edward J. Lawler. San Francisco: Jossey-Bass, Inc., 1981. Pp. xvii + 234. $15.95.

Selwyn W. Becker
University of Chicago

SCENE:	A singles bar
FEMALE,	34, voluptuous, with long wavy brunette hair: Hi, I'm Laura. What's that you're drinking?
MALE,	with one snifter in each hand: This one is cognac; and this one is heavy cream. I use it for a chaser.
FEMALE:	What an innovative way to drink a Brandy Alexander! I have some cognac at home. Why don't we get some cream and go to my place for the evening?
MALE:	Thanks a lot, but no. There are 40 or 50 other ladies here I haven't talked to yet and besides I just performed at a sex clinic because they needed a 70-year-old surrogate.

Samuel B. Bachrach and Edward J. Lawler would say that the female in the dialogue above has no power over the male because he is not dependent on her. He is not dependent because he has many alternatives (40 or 50 other ladies) and he does not ascribe much importance to the projected outcomes of her offer (how many times can the average 70-year-old perform in one day?).

That definition of power and dependence is clearly stated by the authors, as is their intent in writing *Bargaining, Power, Tactics, and Outcomes:* ". . . to examine the determinants and consequences of bargaining power and link bargaining power to the use of particular tactics" (p. x).

The authors criticize some classical theories of bargaining that attempt to predict bargaining outcomes. Since most of the theories explicitly (but perhaps not implicitly) neglect "power" as a construct and are not concerned with tactics, the apparent reason for reviewing this literature is because "it is there."

The authors next assume that "bargaining power is the pivotal construct for a general theory of bargaining" (p. 43) and adopt the social-psychological exchange theories definition of power as dependency of the other. Again, dependency is defined by the number of alternatives available to the party and by the party's commitment to the outcomes (importance ascribed to the outcomes, rather than the magnitude of the outcomes). They then deduce some propositions about how the two dimensions of dependency affect concessions behavior (bargaining toughness) and present data from laboratory bargaining experiments as evidence.

The first hypotheses state that an increase in A's alternatives will decrease A's concessions and will either increase B's concessions or have no effect on B. How A's alternatives affect B depends on the bargainers' images of power. There is no variable in this theory which relates to bargainers' images of power and hence no prediction regarding B's behavior. We are left with the essentially ad hoc finding that an increase in A's alternatives did increase A's toughness. A made fewer concessions, but this had no effect on B.

Next the authors look at how commitment affects concessions. Strangely enough, they operationalized commitment by varying magnitude of the outcomes. They hypothesized that increasing A's commitment will either increase A's concessions or decrease A's concessions, depending on how A views power. Since there is almost no way to defeat this hypothesis, it is almost superfluous to state that it was upheld.

They then turn to the important area of how dependence (the other's punitive capability) affects threat and punishment tactics. Using their dependence theory, they hypothesize that increased punitive capability within the relationship will decrease the frequency of threat and punishment tactics. They contrast their hypotheses to those from the theory of conflict spiral. This theory, based on social-psychological experimentation, permits the conclusion that punitive capabilities lead to threat-counterthreat and punishment-counterpunishment. The experimental data

presented in this book are much more supportive of the theory of conflict spiral than of the authors' own deductions.

At $15.95, this book on bargaining is no bargain.

Futures for a Declining City: Simulations for the Cleveland Area. By Katherine L. Bradbury, Anthony Downs, and Kenneth A. Small. New York: Academic Press, 1981. Pp. xiii + 247. $27.00.

Eugene S. Uyeki
Case Western Reserve University

Futures for a Declining City: Simulations for the Cleveland Area is an important simulation of the future conditions of a prototypical metropolitan area in the United States which is declining in population and employment. It is the detailed case study of an American city which is summarized in chapter 11 of the authors' *Urban Decline and the Future of American Cities* (Washington, D.C.: Brookings, 1982). Cleveland and its SMSA form the example chosen for study because Cleveland is "typical of large industrial areas of the U.S. manufacturing belt" (p. 2). The motivation of Katherine L. Bradbury, Anthony Downs, and Kenneth A. Small is based on the assumption that there "is *some* case to be made for public intervention to slow the loss of activity from large central cities" (p. 4). The base case projection assumes a continuation of present trends and policies. The other projections assume various policies designed to stem the loss of population and employment.

The urban variables on which they focus are (1) employment, (2) population, (3) housing, (4) local government fiscal conditions, and (5) transportation. They project changes in these variables from 1975 to 1990, using various assumptions and analytic techniques. The typical method of the authors is to consider two sources of change for any variable: a change in the variable for the entire United States and a change in the variable for the county under study. They used projections made in 1975 by Northeast Ohio Area Coordinating Agency, the local area-wide planning agency for employment and population, but revised the latter downward because they felt the projections were overoptimistic. Even their reduced population estimates are somewhat larger for the central city (by 0.1%) and the counties (by 0.8%–1.9%) than the results from the 1980 census.

Their projections suggest that, if present trends and policies continue over the 1980–90 period, the city will lose more (or gain less) than the rest of the SMSA. If the city drops in population by 18.2%, the suburbs gain 0.9%; if the city loses 10.8% in households, the suburbs gain 10.4%; if the city loses 17.3% in jobs, the suburbs gain 16.4%; and if the increase in per capita income in the city is 19.9%, the increase in the suburbs is 25.1%.

The most interesting parts of the book are the projections of changes induced by policies addressed to slowing down the losses of jobs and people from the central city. The strategies of intervention are (1) job stimulation, (2) housing rehabilitation, (3) transit improvement, (4) fiscal equalization, (5) suburban growth control, and (6) all-out revitalization—all policies implemented together. Another projection considers the consequences of energy scarcity on locational decisions of employers and households. In addition to discussing the primary effect of any policy, the authors take into account the secondary and induced impacts—that is, indirect relocations generated by the direct efforts to motivate movement into the city of jobs and households or to slow down the loss of these from the city.

Compared with the base case, the all-out revitalization strategy would produce 32,600 more jobs, 25,700 more households, and 64,500 more people over the 1980–90 period. Jobs in the central city would decline by 8.7%, population would decline by 6.9%, and households would increase by 1.1%. This most optimistic of projections suggests that one can slow the decline in jobs by almost two-thirds and that of the population by almost one-half. The single policies have differential impacts on job and household changes. The job stimulus package and the transportation package affect job changes more than household ones. Housing and suburban growth control affect household changes more than job changes. The fiscal equalization policy affects both job and household changes. The all-out strategy does not add up to the total of impacts on job and household changes induced by each of the separate policies because of interaction among them. A chapter considers and dismisses the effect of the high cost of energy as a significant factor in urban development for the next one or two decades.

The authors are aware of the difficulties of quantifying the complexities in urban development as well as of the low probability of the joint occurrence of the various events and activities—private/public and intergovernmental—required to implement the various policies. These problems lead to the conclusion that each of their public policy recommendations would be underimplemented. For example, it is difficult to conceive that one could coordinate, over enough time to make a real difference, a broad-gauged job stimulus package composed of the following elements: training subsidies, loan program for private investment, doubling of the number of CETA jobs, federal priority for location of jobs in central cities, and city assistance with land parcel acquisition and streamlining of legal assistance for business development. Other policies, particularly fiscal equalization and suburban growth controls, require a high degree of city-county cooperation, including consolidation, which has occurred in only a few cases in the United States.

One should not be overly pessimistic, however, for changes are taking place that may help slow the process of decline. Since this book was written, a four-year term has been set for the mayor of Cleveland—a step seen by the authors as important in providing longer-term policy stability

in the central city. Sohio—an energy company—is building its $200 million plus corporate headquarters on one side of Public Square in downtown Cleveland. This will have important multiplier effects in many areas over the long term.

In sum, the authors have written a careful study in the quantitative assessment of the consequences of several policy strategies to stem the decline in jobs and population in the central city of Cleveland. The results for Cleveland can be generalized to other cities in the Midwest and eastern United States which have gone through a similar history of population increase and decline. It is chastening to take a long view about these events. Population and jobs have been moving away from the central cities in the United States since the 1920s and particularly since World War II. More recently, in the 1973–80 period, jobs and population have increased more rapidly in nonmetropolitan counties. In other advanced industrialized countries, there is a shifting of the population and jobs from the core area to the periphery—a similar process of decentralization. The process of decentralization will undoubtedly be hastened by the democratization of computer usage in the home for work, education, and leisure. The present study shows the considerable efforts needed to slow the decline in the central cities which are being sustained by society-wide social forces.

Cities and Suburbs: Urban Life in West Africa. By Margaret Peil. New York: Africana Publishing Co., 1981. Pp. 322. $35.00. (Available from Holmes & Meier.)

William E. Bertrand
Tulane University

As Africa becomes more important in world geopolitics, scholars are showing greater interest in data-based works on the area. Margaret Peil in her ambitious *Cities and Suburbs* has focused ably on one of the least studied phenomena in Africa, comparative urban patterns. Using traditional survey techniques, Peil analyzes data from Banjul and Serekunda in the Gambia; Tema and Ashaiman, in Ghana; and Aba, Abeokuta, Ajegunle, and Kakuri, in Nigeria. The interviews were carried out in the late 1960s and early to mid-1970s. Total sample size was rather small for comparative multisite research, ranging from slightly over 200 to 460, with the mean number of interviews per site being 259.

The volume is organized into two sections: part 1, which focuses on social structure, and part 2, on social interaction. The social-structure section includes, after a general introduction, sections on migration, work, and housing. Migration is discussed, with the finding that migrants have higher qualifications and therefore tend to do better than urban-born counterparts. Peil's conclusion that rural-urban migration will not be slowed by rural development programs is an important one which should

be noted by developmental specialists working in the African region. Peil's statement of explanation in support of the attraction of the city ". . . because the urban standard of living is sufficiently higher than the rural standard . . ." (p. 64) reinforces the idea that migrants are making a rational choice in their move to the cities. The extensive treatment of the "going home" syndrome which appears to be a major difference between African and other developing country urban migrations is well documented, and the liberal inclusion of examples in a case report form throughout the volume enlivens the reading.

A final useful note in the first section has to do with the careful and methodologically important classification of occupations. From my own fieldwork in urban Africa, I consider this to be a major methodological problem, and Peil's attention to it offers an important research tool to the many individuals faced with this problem in their own research.

The second section of the book is dedicated to social interaction patterns and offers little that is novel to students of African society. Social interaction is found to be dependent on housing patterns and occupational clustering, with economic differentiation playing an important role. The predominant feature influencing African urban social life is ". . . that the majority of adults are or have been married and that migrants seldom cut themselves off from their families or kin" (p. 139). This reinforces more traditional influences and their continuation. The finding that sex ratios are tending toward balance contrasts with earlier reports on African urban populations documenting a predominance of males.

A final point worth mention is the emphasis on the so-called personal approach in urban Africa, a correlate of the more traditional patterns documented at all levels. Peil points out that this ". . . makes it almost impossible to eliminate corruption and reposition from the societies" (p. 266). The different African way of doing business has often been commented on, but the problem has seldom been understood as a natural result of social organization. Although this observation is touched on only lightly, I believe that it is an important one.

In a part of the concluding chapter, Peil ventures briefly into developing an urban typology with a focus on clusters of attitudinal scales used in the prior chapters. While she hedges a bit owing to the high levels of unexplained variance in her models, she nevertheless makes the reasonable statement that a metric exists which ". . . corresponds roughly to an increasing level of industrialization; complexity of the labor force and exposure to influences of national development" (p. 267). My major criticism of the book is precisely the lack of such synthesizing efforts. The mix of interests and slightly confusing presentation lead the reader to search for the continuity in the comparative study. This may be due to the difficulties in dealing with multivariate phenomena with relatively small total numbers and many variables. While the statistical analysis appears sound, the limitations of the data may have added to problems in interpretation.

These are relatively minor problems, however, when balanced against

the overall contribution of the book. It provides important descriptive information of the modern era in urban life in four different countries. The logistics of such an effort alone are worthy of our praise. Peil has provided a quality work which should be in every collection related to Africa and will be useful to those in development or comparative urban studies as well as to those area specialists with interests in the specific countries.

The Chinese in America. By Jack Chen. New York: Harper & Row, 1980. Pp. xiv + 274. $15.95.

William Liu
University of Illinois at Chicago

In this lucidly and beautifully written account of Chinese immigrants in America from the 19th century to the present, Jack Chen has done a superb job of casting history into a perspective of broad understanding of nation building combined with a sense of ethnic pride. The emphasis on the Chinese in the 19th century in place of an even treatment of them in the past and present (the post-1965 era) is both obvious and understandable in the light of Chen's role in the mounting of a national exhibition, which began in 1980 and went from city to city through 1981. This exhibition, called the "Chinese in America: 1840–1980," uses artifacts, correspondence, official immigration, and government documents, as well as family and community photographs, to tell the stories of Chinese immigrants and the early days of San Francisco. To understand the structure of the *The Chinese in America,* one would have to follow the design and the development of the national exhibition—they are the twin products of the author's recent work. The common features of both projects are early immigrant histories and a focus on the West Coast. This is a book which should be read by those who are interested in the Chinese in America and those who are interested in California history. Even though both subjects have been written on extensively, Jack Chen still finds some new stories to tell and new ways to tell old stories. Among these stories are accounts of the early California offshore fishing industry (chap. 8) and the fish canneries in Oregon and the Columbia River area (chap. 7).

There are many reasons why an author wishes to write a book. For Jack Chen this work was almost destined from the beginning. The author's birth in Trinidad of immigrant Chinese parentage barred him from becoming a resident immigrant in the United States, although many of his kinsmen who helped build the American West were from the same village as his grandparents and parents. The diaspora of people from a few small villages in Guandong Province was determined as much by the search for economic opportunities as it was by political events on both sides of the Pacific Ocean. Jack Chen spent more than a decade working

for the *Peking Review* (now *Beijing Review*), one of the very few English-language publications in the People's Republic of China. He subsequently lived on the East Coast before he finally settled in San Francisco where he became fascinated by both the country of origin and the host country of many of the San Franciscan Chinese. As a "China man" educated in the West, Jack Chen is, as he says in the beginning of the volume, "a travelling type of man" with a mind that combines hard logic and soft intuition. The book contains arguments supported by detailed anecdotes as well as not so detailed statistical figures.

The book is divided into three broad sections, each progressively treated in less detail than the previous section. The 10 chapters in the first section tell the story of the coming of Chinese to America: how they came; what they did for a living; where they lived; building of the railroad and other kinds of industrial labor participations. Section two, in its seven chapters, deals with the legislative exclusion of the Chinese that was preceded by a pervasive and strong anti-Chinese climate both in California and in Washington, D.C. The discrimination against the Chinese had turned the Chinese communities into exclusion Chinatowns, a unique type of cultural and racial ghetto never before seen in America. Finally, section 3, containing three chapters, deals with the modern-era Chinese and recent Chinese immigrants.

The strength of this book lies, in part, in Jack Chen's use of an insight stemming from a sense of his personal search for roots. Some of the stories told include the author as part of the drama of how immigrants experienced hardships which left them with bittersweet memories. In his work with historical materials, the author projects himself into the past at the expense of much of the more complex and less written about materials on Chinese immigration and the metamorphosis of Chinese communities in the post-1960 era. The author's discussion of these in the last three chapters and the appendix (on Chinese communities in America) is skimpy and based on inadequate use of census materials. Much can be done with the available materials on the new waves of Chinese immigrants. Despite some shortcomings, Chen raises significant questions on the gaps in the Chinese-American story. For example, the description of the decline of the Zhongshan immigrant community in Locke, California (chap. 20), is significant because few authors have touched on this subject previously. Chen wonders how such a community of a cohesive and aspiring group of earlier settlers was allowed to die, without leaving a single clue as to why. In chapter 21, the author states that the large metropolitan Chinatowns, such as those in San Francisco and New York, share the same fate—"to some extent their problems are similar" (p. 234) because "They, too, face problems of conservation, preservation, and development that clearly go beyond themselves . . ." (p. 234). Here, Chen suggests that the survival of urban Chinatowns in metropolitan America is, at least in part, associated with the neglect of the municipal governments. In the past 20 years, decay of inner cities has continued despite large infusions of government money. Chinatowns everywhere are unique in showing an

opposite trend of business prosperity and vitality in spite of government neglect. The real problems lie not in the possibility of the death of a community but in overcrowding, inadequate health and housing facilities, and the total lack of recreational space. More Chinese than before are willing to live in or close to Chinatowns, so the demise of such enclaves is unlikely; it is more likely that there will be changes, but how will these come about?

New Urban Immigrants: The Korean Community in New York. By Illsoo Kim. Princeton, N.J.: Princeton University Press, 1981. Pp. xvi + 329. $26.50.

Peter Park
University of Massachusetts—Amherst

The 1980 U.S. Census counted 354,529 Koreans in the United States, virtually all of them recent immigrants of the past 15 or so years, about 80,000 of whom, according to Illsoo Kim, have settled in the greater New York metropolitan area. *New Urban Immigrants: The Korean Community in New York* explains the emergence of this ethnic subpopulation in the New York area and attempts to understand its internal dynamics. The author places the study in the framework of a community study by asking, at the outset: Do the New York Korean immigrants constitute a community in the sense the term has been understood in the sociological tradition? But, anticipating difficulties, he also asks: Is the traditional concept of community a useful one for understanding the new urban immigrant population? In ostensibly attempting to answer these questions, Kim undertakes a historically sensitive analysis of the political, economic, and cultural factors that have shaped the Korean immigration.

Kim explains the sudden flow of immigration from South Korea, which has been the second-largest source of immigration in the Eastern Hemisphere since 1965, in terms of the familiar concept of "push" and "pull" factors. On the "push" side, he lists such circumstances as: the presence in South Korea of alienated and predominantly Christianized political refugees from the North; urbanization; overpopulation; surplus labor resulting in underemployment and unemployment, especially among the educated urban youth; exploitation and suppression of labor; the emergence of a new entrepreneurial class enriched by a dependent economy; and the political anxiety, felt especially by the newly rich, caused by the constant threat of war between the divided halves of the country. On the "pull" side he mentions: labor shortages in the U.S. medical professions, particularly in urban hospitals; opportunities in small businesses in slum and transitional areas; the lure of affluence propagated by the 40,000 or more U.S. servicemen stationed in South Korea who are a conspicuous advertisement for the virtues of the American way of life.

Kim is at his best when tracing the structural forces creating some of

these push and pull factors, which he places in a historically informed politicoeconomic context. This study thus contains potentially useful material for filling in some of the gaps in the theory of migration in the world-systems perspective because it deals with professional workers and small-business people, two classes often neglected in modern migration theories. Nevertheless, the author is apparently oblivious of such a theoretical perspective and, by failing to be cognizant of it, misses the opportunity to analyze what he calls the push and pull factors of the Korean immigration as two sides of the same coin, joined together and complementing each other by the logic of the capitalist world system.

Kim is less than successful in giving a rounded picture of the New York Korean community because his analysis of the community, which relies heavily on popular media sources, is limited to the more visible activities of small-business owners, medical professionals, churches, secondary associations, ethnic newspapers, "umbrella" organizations of the community, and the Korean Consulate General of New York. The author, by his own account, utilized 20 "informal" informants, interviewed hundreds of immigrants, and conducted participant observation in the community. But the description of the community life that emerges is thin in texture. There is no statistical description of this population—no breakdowns by such standard characteristics as age, occupation, education, etc.—since the study does not employ a survey methodology. In view of the problems associated with obtaining valid information from surveys, especially among immigrant groups, it was perhaps a wise strategy to eschew deceptively objective statistics of this vaguely defined community. But in its place the author should have provided a more interpretively anchored analysis of the life among the immigrants. As it is, the reader finds it difficult to feel the workings of the ethnic solidarity that the author discusses, since there is no phenomenological display of how it manifests itself or fails to do so.

Does the Korean population of New York constitute a community? Kim's answer to this question is oscillating and ambiguous. He identifies, but does not analyze, social factors that tend to produce ethnic solidarity among this population, and other such factors which operate in the opposite direction. Among the latter he includes the supposed character structure of the Korean immigrant which, as he takes pains to show, has been shaped by the conquest-ridden history of Korea and the unstable class structure of the traditional home society. In the final analysis, Kim's prognosis for the emergence of a Korean community is tentative. In any event, he does not feel that the community that is, or is in the process of becoming, can be understood in terms of the traditional concept of community which includes *Gemeinschaft* and territoriality, since these Koreans were already urbanized in Korea and have settled in a dispersed pattern over a large urban area—hence the title of the book. Is, then, the hallowed sociological concept of community a useful one for understanding this group of immigrants, and others like them, who have been created by global capitalism? The author unfortunately fails to rise to the chal-

506

lenge of this question and misses the fine opportunity presented to him to contemplate a fresh way of looking at the phenomenon of urban immigrants that is a fact of life in the modern world.

Professional Work and Marriage: An East-West Comparison. By Marilyn Rueschemeyer. New York: St. Martin's Press, 1981. Pp. 179. $25.00.

Robert S. Weiss
University of Massachusetts—Boston

This slim book is interesting for the questions it raises rather than for the findings it presents. But the themes it tackles are large ones, and even though what it has to say is more nearly suggestive than conclusive, *Professional Work and Marriage: An East-West Comparison* is well worth reading.

Marilyn Rueschemeyer interviewed a total of 60 respondents living in four places—the Boston-Providence region, West and East Berlin, and Israel. The Americans and the East Berliners were interviewed about their current lives. Those in West Berlin and in Israel were former Soviet citizens, most of them Jews, whom she interviewed about their lives before their emigrations. Rueschemeyer groups her interviews under three headings: American single-career families; American dual-career families; and socialist society dual-career families. The cell for socialist society single-career families is empty; apparently, at the occupational level of Rueschemeyer's respondents, women are as much expected to have careers as are men.

Rueschemeyer characterizes the occupational stratum of her respondents as "professional." It is important to note that, by this, she does not mean that her respondents are doctors or lawyers or members of the clergy, although some of them are, but rather that they are in upper-middle-class lines of work. They include engineers, businessmen, and quite a few academics.

Rueschemeyer finds that her American men work very long hours. Several are said to try to work all the time; virtually all work weekends and evenings. As one wife said of her husband, "He seems always to be running."

There is ample justification at the beginning of the career for this: the need to get established. But the work-centeredness does not let up, even after the man has built his practice or won his promotion or obtained his tenure. Not only do they fear losing their job; even more, they fear not being respected by their colleagues. In addition, there is competition for rewards: funding, invitations, recognition. Finally, although Rueschemeyer does not make much of this, there is the pure joy of the work, of engagement and success.

No wonder that the wives who have not themselves bought into this

507

are perplexed. They think their husbands spend too much time away from home and give too much energy to work. When the husbands are at home, they are too often preoccupied or just exhausted. While the husbands try to be good fathers, they do their fathering in brief bursts and then they are back at work. The women understand none of this. Still, the women are not without their ambivalences; the men's success brings them income, security, and position in the community.

Into this situation, for several of Rueschemeyer's respondents, came the Women's Movement. Rueschemeyer says: "By far the most significant change in the marital relationship came from the influence of the Women's Movement, which drastically affected eight of the ten women in their view of their situation . . ." (p. 43). With resolve gained from consciousness-raising groups, these women developed career-centered lives of their own.

Now, in the resulting dual-career families, there were two people, each devoting the bulk of time and energy to work. How could they manage this and yet meet their responsibilities? One of them, most likely the wife, might accept having a second-level career and take primary responsibility for home and children, or husband and wife might share responsibility for home and children and make whatever sacrifices of work effort are required. There is still another possibility: husband and wife could each reduce their commitment to the home and children and, perhaps, to the marriage as well.

Most dual-career couples did cut down on child care, using substitute caretakers if the children were small, leaving the children on their own if they were older. They also cut down on entertaining—and, possibly, on the number of their shared friends. Most found, too, that their marriages were emotionally stressed and thinned. They valued vacation times when they could go off with their mates.

Rueschemeyer wanted to find that it was the competitiveness and insecurity of American life that was responsible for all this. Her interviews with veterans of socialist society surprised her. Despite having grown up within an ideology of cooperativeness, her socialist respondents displayed as much work-centeredness as did her American respondents. There were differences, to be sure; in socialist society there was concern about retaining the respect of the work collective rather than of some less well-specified group of colleagues; in socialist society, too, there was the constant threat of being identified as politically unreliable. But the fundamental dilemma remained: work was insatiable in its demands, and how was a family life to be maintained if both husband and wife were forever responding to them?

True, socialist society maintained as an ideological tenet that equality to men in the workplace would free women from tyranny at home. But the result did not seem to be a couple whose love had been cleansed of the resentments of a master-servant relationship. Rather, it seemed to be a couple whose love had disappeared in the course of each member's striving for success.

It appears from Rueschemeyer's materials that work such as that performed by managers and professionals is engaging precisely because success in the work is a basis for self-regard and dignity and, in addition, the chief means for eliciting the regard of others. Because "professional" work has these implications, and because it leaves to the person doing it the decision of how much is to be done and how well, its demands are without limit. The one-career marriage maintains a family life despite the insatiability of the work by keeping one member free of its entanglements. This approach, however, requires that the work-free member perform the boring and yet wearing tasks of keeping the house in order and the children cared for, while the other member of the marriage has the excitement of work and the reassurance of success. The two-career family remedies this inequity, but whether it provides as satisfactory a base for marriage and a family is yet to be seen. Rueschemeyer suggests that attitudes toward work may first have to undergo change. I agree but do not think it will happen.

Rueschemeyer raises still other issues in her book. They include the substantive issues of the styles of marital relationship that are fostered by the intense work commitment of one or both members of the marriage and the particular forms of marital stress that result from work-centeredness. They include, in addition, the methodological issues of how much can be learned from interviews with small samples chosen opportunistically and what confidence should be placed in retrospective accounts. It is a merit of the book that it brings these additional issues sharply into focus.

Generations Apart: Adult Hostility to Youth. By Leon Sheleff. New York: McGraw-Hill Book Co., 1981. Pp. xii + 351. $18.95.

Ilene Philipson
University of California, Berkeley

In *Generations Apart: Adult Hostility to Youth,* Leon Sheleff reviews an astonishing number of books and articles to demonstrate a single point: parents and adults are and always have been hostile toward children and youth. This hostility, according to the author, is a transhistorical, cross-cultural, and inherent truth that has remained obscure only because it is parents who record, theorize, and transmit information about parent-child relations. "[P]arents have a built-in antagonism toward their progeny; . . . their attitude toward them may range from ignorance and indifference, through denial and contempt, to open attack and ultimate destruction" (p. 37).

Sheleff's purpose in writing this book is to demonstrate how " 'adults make so much trouble for youth' " (p. 10), and he does this quite successfully. *Generations Apart* is a veritable catalog of examples of the adult

hostility which is expressed through familial, medical, educational, and judicial institutions. This work also provides an extensive review of the literature on parent-child relations and explores the inability or refusal of a great many authors to recognize antagonism toward children—in their own work, in history, in psychoanalysis and social theory, and in literature. Those few authors who recognize or discuss this hostility are celebrated, while those who do not are criticized.

The one thinker who is criticized the most for masking and denying parental hostility is Freud. Sheleff believes that Freud's conceptualization of the Oedipus complex has had the greatest influence in obscuring antagonism toward children. By emphasizing children's murderous wishes toward their parents, Freud obfuscated the more salient and portentous parental hatred of children.

While the author may be correct in criticizing Freud and his followers for relying too heavily on Oedipal explanations of psychic and social experience and activity, I believe Sheleff commits the same error in attempting to explain the most diverse phenomena involving parents and children in terms of innate adult hostility toward youth. For Sheleff, everything from infanticide to arranged marriage, from the use of wet nurses to child labor, transcends historical and cultural specificity, loses any social, political, or economic meaning, in order to be interpreted exclusively in terms of the common denominator of parental hostility.

In this sense, *Generations Apart* does not provide a sociological analysis. While it is necessary for the author to examine works of sociology, history, anthropology, etc., to chronicle adult antagonism toward children, the explanation of *why* such hostility exists is purely biological. Through the first three-quarters of the book Sheleff fails to mention why parents might harbor such antagonism toward their offspring. It is not until page 242 that we discover that the birth of a child signals "the beginning of decline, of eventual mortality" for the parent, and that this "constitutes a dilemma from which there is no escape, a dilemma for which there is no solution. It is a structural factor inherent in the very nature of generational contact, irrespective of any cultural overtones that mitigate this fundamental fact." Thus we are left with the assertion of a supposedly immutable biological reality that parents, as social actors, are incapable of fundamentally altering.

The questions of why a child's birth automatically triggers realization of a parent's mortality, of why this realization necessitates parental hostility, and of why this situation is exempt from social influence are not seriously addressed in *Generations Apart*. For it is the author's purpose only to document the fact that hostility does exist: "For what is needed above all is an awareness of tensions and hostilities that we are too often prone to ignore" (p. 332).

While Leon Sheleff certainly fulfills his purpose, he contributes little of theoretical import and nothing new either from studying history or from analyzing the present empirically. His work remains a well-researched, comprehensive, and extremely lengthy book review of other

authors' recognition, or, more frequently, lack of recognition of adult hostility toward youth.

Illegitimacy. By Jenny Teichman. Ithaca, N.Y.: Cornell University Press, 1982. Pp. viii + 200. $18.50.

Ann Oakley
National Perinatal Epidemiology Unit, Oxford

Philosophy in the English-speaking world, says Jenny Teichman in *Illegitimacy,* tends to be pure philosophy. What is pure philosophy? It is the examination of *general* notions which are not specific to any branch of human social life. In speculating on such matters as cause, time, and space, the philosopher is not concerned with their variety of social and personal meanings but only with whatever essential meanings they can be said to have independent of these. To the sociologist, therefore, pure philosophy is itself essentially uninteresting, an arid exercise.

In asking what the idea of illegitimacy means, Teichman is not, however, confining herself within such limits. She is, instead, practising an art (science?) named applied philosophy which, as its title suggests, "applies" the rigors of philosophical thinking to selected areas of human social behavior, and which, accordingly, requires some knowledge of the working of actual human institutions. The idea of illegitimacy is, in Teichman's words, "situated at a multidimensional interface of human institutions" (p. 10). Understanding the concept demands a confrontation with what human beings have done with kinship, lineage, property, law, morality, religion, birth control, and reproduction. This multi-institutional aspect makes "unpacking" the idea of illegitimacy both particularly interesting and especially difficult.

It helps, of course, to believe that ultimately it *is* possible to arrive at a more-or-less satisfactory universal definition of legitimacy and illegitimacy, and this is the chimera Teichman resolutely chases throughout this book. She asks, for example, whether marriage is a necessary and sufficient condition for the existence of the ideas of legitimacy and illegitimacy in all human societies, thereby entangling herself in the debate about whether the institution of marriage can itself be said to be in any sense universal. Her answer is that, since marriage can only with enormous difficulty be said to exist in the same way in all cultures, it is merely a "paradigmatic precondition," not a necessary or a sufficient condition for illegitimacy. Perhaps the most that can be said is that an illegitimate child is one whose existence is the result of an unsanctioned sexual act. Which sexual acts are, or are not, sanctioned, is then left to the vagaries of national custom to decide.

As a matter of fact, some of the least satisfactory customs of our own (Euro-American industrialized) culture concern the sanctioning of sex-

511

uality in marriage and the legal status of children born in wedlock. It is assumed, for instance, that the right of a woman to refuse intercourse with a man cannot be violated in marriage, because that right is precisely lost to a woman upon marriage. It is also assumed that the children of a married woman are her husband's children, the sole exception to this rule apparently being that a child conceived in one marriage and born in another is presumed to belong to the first husband, except in West Germany where the second husband legally acquires the child as his own. Yet there is evidence from samples of English families in whom, for one reason or another, blood tests have been done, that a significant proportion of such children assumed to be legitimate are not.

As a historical and cross-cultural survey of the many imaginative strategies human beings have employed to distinguish between legitimate and illegitimate children, Teichman's book is a most valuable contribution. One of the outstanding features of illegitimacy is the disability that has traditionally attached to being illegitimate in most cultures, a disability well reflected in such diverse fields as literature and infant mortality statistics. The table of legitimate and illegitimate infant mortality in England and Norway (p. 105) contains an unfortunate but appropriate misprint—infant *morality* is printed instead of infant *mortality*. It is perhaps not difficult to understand why a society bothered about the legitimate inheritance of property should want to punish illegitimate people. It is, however, quite a problem to predict what will happen to illegitimacy in the future when, in some countries, the proportion of children born outside marriage is rising epidemically (37% in Sweden in 1979). Most suitably, Teichman does not indulge in crystal-gazing, contenting herself with a documentation of the various attempts that have been made to abolish the legitimate/illegitimate distinction.

It is something of a condemnation of the myopic vision of sociologists that a book like this should be written by a philosopher. Placing illegitimacy in its broadest possible context, Teichman takes it well outside the "social problem" perspective which has bedeviled its treatment by sociology. All too frequently, sociologists have naively adopted the line that illegitimacy, being "abnormal," needs to be explained, whereas legitimacy, being "normal," does not. In this manner, sociology has confirmed, rather than elucidated, the moral disabilities of illegitimate children and their mothers—disabilities which spring ineluctably from human institutions and not from natural facts.

Chicago

Family Life and School Achievement
Why Poor Black Children Succeed or Fail
Reginald M. Clark
Clark offers intimate portraits of ten black households—from both ends of the economic spectrum, in one- and two- parent homes, with high and low achievers—to show that income level and family structure do *not* inflexibly dictate school success or failure. Emphasizing that the key factor for academic success is the *quality* of family life, he outlines how we can facilitate change in urban families.
Cloth $22.50   264 pages

How Schools Work
Rebecca Barr and Robert Dreeben
With Nonglak Wiratchai
This book explores the ways different parts of an educational system and the resources they utilize contribute to learning. The authors not only document learning effects, they also show where they are produced in the system, where they are not, and why. *How Schools Work* makes a major contribution to unraveling the conceptual tangle of how learning is achieved.
Cloth $22.50   208 pages

New in paper

The Erosion of Childhood
Valerie Polakow Suransky
The Erosion of Childhood should be required reading for all concerned persons; and all persons are concerned with children and the future, whether admittedly or not....The information and questions contained in the book could lead to a substantial revision of our child care policy and practices."—*The Journal for Special Educators*
;7.95   236 pages

Damaged Parents
An Anatomy of Child Neglect
Norman A. Polansky, Mary Ann Chalmers, Elizabeth Werthan Buttenwieser, and David P. Williams
Although less dramatic and visible than outright child abuse, child neglect occurs five times more frequently and contributes to widespread social problems. In this book, based on fifteen years of research, Polansky and his coauthors report on two systematic efforts to define, measure, classify, and understand child neglect. "Polansky and his colleagues have produced a truly remarkable book."—Thomas M. Young, *Social Service Review*
$7.95   284 pages

THE UNIVERSITY OF CHICAGO PRESS
5801 South Ellis Avenue   Chicago, IL   60637

CONTENTS

Review Essay

Book Reviews

IN THIS ISSUE

MARK COLVIN is a graduate student in sociology at the University of Colorado, Boulder. His areas of interest include criminology, political economy, and social stratification. He has published articles on the 1980 New Mexico prison riot and is currently conducting research on the political economy of imprisonment and welfare policies in the United States.

JOHN PAULY is a doctoral student at the University of Colorado, Boulder. His doctoral thesis is a structural-Marxian analysis of class relations and social control in television news production, based on participant observation of local news operations and content analyses of local television newscasts.

TRAVIS HIRSCHI is professor of sociology at the University of Arizona. He is currently working on the implications for crime theory of factors (such as age) that appear to have a direct impact on crime. His recent work includes a chapter on crime and the family in James Q. Wilson's *Crime and Public Policy*.

MICHAEL GOTTFREDSON received his Ph.D. in criminal justice from the State University of New York at Albany in 1976. He recently completed a study of policy guidelines for bail in the Philadelphia Municipal Courts (with John Goldkamp). He is associate professor of sociology at the University of Illinois at Urbana-Champaign.

ROBERT L. KAUFMAN is assistant professor of sociology, on leave from the University of Texas at Austin to fill a visiting position at the University of Utah. His research specialities are labor market segmentation and statistics. He is continuing his work on labor market segmentation and minority opportunities and is also studying issues in multivariate cluster analysis.

DONALD J. TREIMAN is professor of sociology at the University of California, Los Angeles. This paper is part of a large-scale cross-national comparative study of social mobility and the process of status attainment in which he has been engaged for some years. He is also involved in policy-related work on earnings differences between men and women.

PATRICIA A. ROOS is assistant professor of sociology at the State University of New York at Stony Brook. Her major research interest has been the cross-cultural study of gender and marital differences in occupational attainment. Other areas of interest include institutionalized mechanisms contributing to occupational segregation by sex and ethnic differences in earnings attainment.

R. ROBERT HUCKFELDT is a political scientist and assistant professor in the Department of Government at the University of Notre Dame. His research interests revolve around the consequences of social contexts for the political behavior of groups and individuals. He is the author of several articles on politics and contexts, and a coauthor (with C. W. Kohfeld and Thomas W. Likens) of *Dynamic Modeling: An Introduction*.

CATHERINE E. ROSS is assistant professor in the Department of Sociology and the College of Medicine at the University of Illinois at Urbana-Champaign. She is interested in the social and cultural determinants of physical and mental health, and her current research focuses on the effects of powerlessness and normlessness on psychological well-being.

JOHN MIROWSKY is assistant professor in the Department of Sociology and the College of Medicine at the University of Illinois at Urbana-Champaign. His current research is on the effect of perceptions of powerlessness and inequity on mistrust, paranoid beliefs, and distress.

PATRICIA ULBRICH is a Ph.D. candidate in the Department of Sociology, University of Illinois at Urbana-Champaign, and an assistant professor at the University of Miami. Her current research examines social factors contributing to the similarities and differences in depression among men and women.

Information for Contributors

Editorial Procedures: All papers deemed appropriate for the *AJS* are sent out anonymously to readers. The editors rely heavily on the judgments of those readers, although they are not bound by them. To protect anonymity, only the title should appear on the manuscript. Attach a cover page with the title of the manuscript, your name, and your affiliation. All identifying references and footnotes should appear on a separate page. Manuscripts are accepted for publication subject to nonsubstantive editing with the understanding that the *AJS* has the right of first publication. Submission of a manuscript to the *AJS* is taken to indicate the author's commitment to publish in this *Journal.* No paper known to be under consideration by any other journal will be reviewed by the *AJS*. Upon publication, all rights, including subsidiary rights, are owned by the University of Chicago Press. The author retains the right to use his article without charge in any book of which he is the author or editor after it has appeared in the *Journal.*

Manuscript Acceptance Policy: While it is our policy to require the assignment of copyright on most journal articles and review essays, we do not usually request assignment of copyright for other contributions. Although the copyright to such a contribution may remain with the author, it is understood that, in return for publication, the journal has the non-exclusive right to publish the contribution and the continuing right, without limit, to include the contribution as part of any reprinting of the issue and/or volume of the journal in which the contribution first appeared by any means and in any format, including computer-assisted storage and readout, in which the issue and/or volume may be reproduced by the publisher or by its licensed agencies.

Preparation of Copy

1. Type *all* copy—including indented matter, footnotes, and references—double-spaced, allowing generous margins at top, bottom, and sides of page. Please do not justify right-hand margins. Number all footnotes, including acknowledgment footnote.

2. Type each table on a separate page. Refer to each table in numerical order in the text. Prepare tables without vertical lines.

3. Clarify all symbols with words in the margin of the manuscript. Encircle these and other explanatory notes not intended for printing.

4. Draw figures on white paper with india ink. Original or glossy print of each figure will be required if manuscript is accepted.

5. Include an abstract of not more than 100 words summarizing the findings.

6. *Four copies* of the manuscript and abstract, on white paper, must be submitted. Manuscripts should be addressed to the Editor of the AMERICAN JOURNAL OF SOCIOLOGY, 1130 East 59th Street, The University of Chicago, Chicago, Illinois 60637. *Manuscripts are not returned to authors.* Enclose

a self-addressed stamped postcard, containing the manuscript title, to obtain acknowledgment of manuscript receipt. No other acknowledgment will be sent.

7. Each article or research note submitted to the *AJS* must be accompanied by a check or money order for $10.00, payable to The University of Chicago Press in U.S. currency or its equivalent by postal or express money order or bank draft. Papers will not be reviewed until this fee has been paid. The submission fee will be waived for papers solely by student authors. Submissions from students should be accompanied by a faculty member's certification of their student status. Citizens of countries with restrictions on the export of U.S. dollars may request waiver of the submission fee.

Format of References in Text: All references to monographs, articles, and statistical sources are to be identified at the appropriate point in the text by last name of author, year of publication, and pagination where appropriate, all within parentheses. Footnotes are to be used only for substantive observations. Specify subsequent citations of the same source in the same way as the first one; do not use "ibid.," "op. cit.," or "loc. cit."

1. When author's name is in the text: Duncan (1959). When author's name is not in text: (Gouldner 1963).

2. Pagination follows year of publication: (Lipset 1964, pp. 61–65).

3. For more than three authors, use "et al." For institutional authorship, supply minimum identification from the beginning of the complete citation: (U.S. Bureau of Census 1963, p. 117).

4. With more than one reference to an author in the same year, distinguish them by use of letters (*a, b*) attached to the year of publication: (1965*a*).

5. Enclose a series of references with a single pair of parentheses, separated by semicolons.

Format of References: List all items alphabetically by author (providing the full list of multiple authors) and, within author(s), by year of publication, in an appendix titled "References." Examples follow:

Davis, K. 1963*a*. "Social Demography." Pp. 124–37 in *The Behavioral Sciences Today*, edited by Bernard Berelson. New York: Basic.
———. 1963*b*. "The Theory of Change and Response in Modern Demographic History." *Population Index* 29:345–66.
Goode, W. J. 1967. "The Protection of the Inept." *American Sociological Review* 32:5–19.
Moore, Wilbert E., and Arnold S. Feldman. 1960. *Labor Commitment and Social Change in Developing Areas*. New York: Social Science Research Council.
Sanford, Nevitt, ed. 1962. *The American College*. New York: Wiley.
Weber, Max. (1921) 1968. "Society's Problems." Pp. 12–16 in *Economy and Society*, edited by Guenther Roth and Claus Wittich. New York: Bedminster.

A Critique of Criminology: Toward an Integrated Structural-Marxist Theory of Delinquency Production[1]

Mark Colvin and John Pauly
University of Colorado, Boulder

Juvenile delinquency is understood as a latent outcome of the social reproduction process of capitalism. A structural-Marxist understanding of contradictory social relations that contour life in modern capitalism allows us to reinterpret and apply the insights from various criminological theories in building a new theoretical approach. A review and a critique of the major theoretical perspectives in criminology are provided, and a structural-Marxist perspective on the larger social structural context is explicated. A new integrated theory of delinquency, grounded in Marxian insights on the role of productive relations in shaping other social relations, is developed and supported with empirical evidence from research representing a broad range of theoretical perspectives.

We attempt in this paper to present a comprehensive theoretical approach to understanding the social production of serious patterned delinquent behavior, defined as repeated engagement of a juvenile in the FBI's Part One Index crimes (Elliot and Ageton 1980). We review past theoretical efforts in criminology in order to reinterpret and build on them in developing a structural-Marxist approach to delinquency. Our approach can be considered "structural-Marxist" insofar as its analytical starting point is the objective structure of social relations grounded in the process of material production under capitalism (Appelbaum 1979; Blackburn 1972; Burris 1979; Gimenez 1982; Godelier 1972).

Our specific concern is with structures of control that solicit and compel certain types of behavior from individuals and shape an ideological orientation for the individual in relation to the agents and apparatuses of

social control. We focus first on social control structures at workplaces and then on those within families, schools, and peer groups. We see these various structures of control as interconnected in a process of social reproduction, contoured by class and production relations of capitalism.

Our analysis borrows heavily from Etzioni's (1970) compliance theory, which postulates a relationship between the type of power employed for control (normative, remunerative, or coercive) and the type of subordinate's ideological involvement in the compliance relationship (moral, calculative, or alienative). Going beyond Etzioni, we may see that the type of control structure in which an individual is immersed arises from: (1) objective class relations, the structure of which defines the extant forms of compliance as well as specific hierarchical positions in compliance structures; and (2) the individual's life-cycle encounters with these institutionalized, class-contoured compliance structures. As cumulative learning processes, these encounters produce both conventional and delinquent behavior. The socialization process can be conceptualized as cumulative experiences arising from such compliance encounters and producing distinct paths of individual socialization. The subjective experiences acquired in these encounters of the individual with compliance structures largely determine the individual's orientation toward authority. Such an ideological orientation can be considered an initial interpersonal "bond" (Hirschi 1969) to specific authorities that, through patterning in repeated compliance situations, is generalized to the social order.

From the point of view of the individual, encounters with compliance structures and the formation of interpersonal bonds begin with the family in parent-child relations. And, with the basic structure of the social order assumed as given, this is where conventional criminology commences its examination of delinquency. However, there is compelling evidence (e.g., Kohn 1977) that family compliance structures are class differentiated and that parent-child relations are profoundly shaped by parents' encounters with workplace compliance structures.

The relations of workplace control, which take various class-related forms under capitalism, shape the consciousness and behavior of parents who repeatedly produce and reproduce control relations with children. The specific nature of these early relations of control, which define the socialization process, strongly influences the type of control structure the child will encounter when entering school; differentiated control structures within schools have been designed for the various labor needs of capitalist industry (Bowles and Gintis 1976; Carnoy and Levin 1976; Noble 1977; Spring 1972).

But much socialization takes place beyond the reach of parents and official agents of socialization. Just as "informal" group organizations of workers emerge within bureaucratic industry, so do peer associations

emerge in the encounters of juveniles with bureaucratic community and educational institutions. Such peer groups develop control relations of their own. The child does not arbitrarily enter a peer group but selects, and is selected for entry, on the basis of previous socialization experiences. Peer groups, through their various structures of control, can reinforce either conventional or delinquent behavior.

The more coercive the control relations encountered in these various socialization contexts tend to be, the more negative or alienated will be the individual's ideological bond and the more likely is the individual to engage in serious, patterned delinquency. We agree with Etzioni that coercive controls create an alienative orientation toward authority. They create negative ideological bonds and open the individual to entry into peer associations that reinforce patterned delinquent behavior (Elliot, Ageton, and Canter 1979).

This entire process of delinquency production is comprehended as a latent outcome of the reproduction of capitalist relations of production and the class structure. We present a class-based theory in that class differences in serious, patterned delinquency are predicted. We recognize the controversy surrounding class-linked predictions of delinquent behavior (Braithwaite 1981; Tittle, Villemez, and Smith 1978). Many of the problems of class-delinquency studies spring from an inadequate conceptualization of objective class categories and from techniques for the self-reporting of delinquency that fail to tap significant differences in frequency and seriousness of delinquent behavior. Recently, Marxist social scientists have begun reconceptualizing the class structure (Edwards 1979; Wright 1978, 1979); these reconceptualizations will be incorporated into our approach. In addition, more sophisticated delinquency self-report techniques, which tap differences in seriousness and frequency of delinquent acts, are beginning to reestablish the association between class and delinquency, at least for more serious, patterned delinquent behavior (Elliot and Ageton 1980).

Many of our theoretical insights are drawn from current criminological theories. Before we can explicate further our structural-Marxist approach,[2] it is necessary to review critically the main theoretical currents in criminology. After reviewing criminological theories, we will return to a discussion of the structural-Marxist approach and attempt to support

---

[2] In this paper we confine our discussion to the processes by which groups of individuals come to manifest a specific range of behaviors that have been declared illegal. The processes by which these behaviors are defined as illegal are beyond the scope of the present paper; this important question has been dealt with by other researchers (Chambliss 1964, 1969; Humphries and Greenberg 1981; Spitzer 1979). While we limit discussion in this paper to patterned delinquent behavior, the structural-Marxist approach has obvious implications for explaining other criminal behavior patterns, such as corporate and white-collar crime, which require further development in future theoretical endeavors.

it with empirical evidence from studies that use a variety of theoretical approaches.

THEORETICAL CURRENTS IN CONTEMPORARY CRIMINOLOGY

The major theoretical currents in criminology today are learning theory, strain theory, control theory, labeling theory, conflict theory, radical criminology, and, most recently, integrated theory. Any theory must be judged by its internal consistency and by its ability to explain a broad range of empirical phenomena. The current theories provide a limited range of explanation and insight on which we attempt to expand by means of our approach.

Learning theory starts with the notion that criminal behavior, like any other behavior, is learned. As formulated by Sutherland (1947), the theory postulates that "one commits criminal acts because his accepted 'definitions' of the law as something to violate are in 'excess' of his accepted definitions of the law as something that can, must or should be obeyed" (quoted in Akers 1977, p. 40). The definitions favorable or unfavorable to criminal law violations result from "differential association" with situations in which these definitions are learned. It is not simply a matter of quantitative association with situations that provide differing definitions favorable or unfavorable to law violation. For, while frequency and duration of associations are important, Sutherland also emphasizes the priority and intensity of the associations, intensity being the emotional attachment to actors in the learning situation. Also, it is not necessary that the associations favorable to criminal law violation be with persons who are actually engaged in criminal behavior; it is only necessary that definitions, or normative evaluations, favorable to such behavior be present for the individual to learn positive attitudes toward criminal law violation. The law violator also must learn the techniques for committing crimes; these will vary with the complexity of the criminal behavior involved. Here association with individuals experienced in a particular criminal behavior may be necessary. The learning of motives and techniques takes place within intimate personal groups.

Sutherland developed the basic scheme for describing a process of socialization toward delinquency. Sutherland's theory has been thoroughly explicated by Cressey (in Sutherland and Cressey 1978) and has been expanded on by Akers (1977) and Burgess and Akers (1966) with conceptualizations from operant conditioning theory (Skinner 1953; Whaley and Malott 1969). Learning theory is limited to the description of the learning process and does not situate this process in a larger social structural context that would give it specific form and probabilities for patterns of content. Akers (1977, p. 64) states that "social structure is an arrange-

ment of sets and schedules of reinforcement contingencies." This insightful specification of Sutherland's thesis nevertheless begs several questions: Exactly how are these reinforcement contingencies differentially arranged for different social arenas and strata, why are they thus arranged, and how and why do these arrangements and contingencies change over time?

Rewards and punishments associated with the learning process are both quantitatively and qualitatively conditioned by the larger social structural context in which they occur. The ways in which learning occurs and behavior is shaped thus depend not merely on immediate rewards and punishments but also on the social structure shaping the reinforcement capacities and patterns, making different rewards and punishments more or less available for persons in different social positions. Accordingly, in order to explain more fully what behavior patterns are learned, and how and why they are learned, the social scientist has an obligation to explore the larger political-economic structure in which learning occurs.

Strain theory attempts to address some of the contextual shortcomings of learning theory. Merton (1938) postulates the existence of a social structural disjuncture between socially desired goals (e.g., material success) and the means for achieving those goals. The disjuncture between goals and means produces a social-psychological strain within the individual, who is then forced to escape the strain through various adaptations. Merton delineated four such adaptations: ritualist, innovator, retreatist, and rebel.

With insights from Sutherland's learning theory, strain theory was further developed by Cloward and Ohlin (1960), who qualified Merton's assumptions regarding homogeneous societal goals and posited that goals and means disjunctures applied not only to legitimate opportunity structures but also to illegitimate opportunity structures. For instance, the "innovator" does not automatically become a thief simply because he is blocked in the legitimate opportunity structure. This individual must also be admitted to, and learn behavior appropriate for, an illegitimate opportunity structure organized around theft. Expanding on Sutherland and Cressey's (1960) notion of "differential social organization," Cloward and Ohlin posit the differential distribution of legitimate and illegitimate opportunity structures which correspond to differential learning experiences. Although, following Durkheim's notion of anomie, they make reference to industrialization, Cloward and Ohlin fall short of specifically explaining the social structural roots of this differential distribution of opportunity structures and the functional reasons for its persistence. They do state that delinquent subcultural orientations (most notably the "criminal" subculture, oriented around theft, and the "conflict" subculture, oriented around violence) appear to vary by availability of opportunities shaped by the differential distribution of legitimate power and material resources.

Strain theory is based on the assumption that goals which are desirable by middle-class standards are shared throughout society. Some researchers support this contention (Cohen 1955; Erlanger 1980; Jessor et al. 1968; Maccoby, Johnson, and Church 1958); others argue that distinct subcultures preclude such homogeneous values (Kitsuse and Dietrick 1959; Lemert 1972; Miller 1958). This issue is open to debate until further information is available on the specification and interpenetration of the ideological structures of different social classes (Laclau 1979; Rodman 1968; Therborn 1980). We suspect that, if the static, mechanistic perspectives on class structure and cultural formations (emphasizing norms and values) were replaced by a more dynamic, dialectical perspective, the main elements of strain theory would be fully reconcilable with subcultural approaches, which view strain as an outcome of contradictory, dynamic *processes* of cultural and class conflict and accommodation (see Simpson and Yinger 1965). As this unsettled controversy suggests, strain theory stops short of attempting to explain the origin of the expectations and discrepant opportunities. The theory merely asserts a goals-means disjuncture as an integral feature of "industrial society." This silence points, once again, to the need for a more precise conceptualization of the overall structural context that shapes expectations, structures of opportunity, and the correspondence between them.

Control theories focus on the factors that insulate individuals from delinquency rather than on those which propel individuals into delinquency. In addition to other contributors to the development of control theory (e.g., Briar and Piliavan 1970; Matza 1964; Nye 1958; Reckless and Dinitz 1970; Reiss 1951; Toby 1957), Hirschi (1969) has elaborated and refined control theory in the direction that has received the most empirical support (Hepburn 1976; Hindelang 1973; Johnson 1979). Hirschi focuses on four dimensions of the "bond" that gives an individual "a stake in conformity" to the "conventional order": (1) *attachment,* the strength of ties to parents, teachers, and peers as primary socialization agents; (2) *involvement,* the degree of engagement in conventional activities that restrict the time for delinquent activities; (3) *commitment,* the devotion to conformist conduct; and (4) *belief,* the strength of conformist attitudes. Hirschi's central argument is that insofar as juveniles' bonds to the conventional order are severed, they are "freed" to commit delinquent acts.

Control theory can be criticized for not specifying exactly how bonds fail to develop or how they may be severed. This theoretical silence is the consequence of control theory's almost exclusive focus on the bonded or unbonded individual while it takes the structure of authority and actions of social control agents as an unexamined given. In our perspective, bonds arise out of the *relation* between individuals and authorities.

Both sides of this relationship must be comprehended in order to under-
stand the bonding process. Bonds are more clearly conceptualized as
ideological orientations toward authority that are learned during the in-
dividual's interactions with agents of socialization (Akers 1977; Conger
1976) in specifically structured patterns of control. The relations of control
in specific contexts are composed of relations between the power wielded
by authorities and the ideological involvement of subordinates (Etzioni
1970). Involvement, in Etzioni's terminology, is an ideological orientation
that corresponds to "bond" as Hirschi uses that term (Thomas and Hyman
1978). Hirschi tends to conceptualize the bond in the quantitative terms
of strong versus weak; the reconceptualization of bonding, following Et-
zioni's notion of involvement, allows us to see qualitative differences in
bonding that correspond to specific types of power utilized by authorities
in specific, structurally defined situations of compliance.[3]

Hirschi's use of quantitative distinctions for bonds derives from his
conceptualization of "society" as a "conventional order" to which an in-
dividual is either "weakly" or "strongly" bonded. The uncritical notion
of a "conventional order," shared by much of conventional criminology,
began to be criticized in the late 1960s by the labeling, conflict, and radical
schools of criminology.

Labeling theory rose to prominence in the turbulent 1960s concurrently
with Hirschi's resurrection of control theory. The former is in many ways
the politico-ideological obverse of the latter. Confronted with a political
climate in which the conventional order was being pervasively called into
question, conventional authority was defensively sanctified by control
theorists and offensively vilified by labeling theorists. Both were bolstered
by the empirical findings of early self-report studies that showed delin-
quent activity to be fairly evenly distributed across races and classes.
Given the ideological predispositions of control theorists, the early self-
report data allowed them to remain oblivious to the class bias of their
bonding concepts and to concentrate on why delinquency proliferates if
deprivation is not the cause ("It must be because of deficient moral up-
bringing"). Given different ideological predispositions, the same self-re-
port findings allowed labeling theorists to remain oblivious to the
determination of delinquent behavior and to concentrate on why certain
groups are officially processed disproportionately if "deviant" behaviors
are evenly distributed ("It must be due to discriminating authority").

---

[3] Etzioni's typologies of involvement (or bonding) can be quantified, with negative bonds
equaling alienated involvement, intermediate bonds equaling calculative involvement, and
positive bonds equaling moral involvement. This conception eliminates the untenable notion
of "unbonded" individuals, who are somehow freed from structural constraints to enter a
Hobbesian state of nature. Instead, bonds of all types are seen as the result of structured
relations of power and subordination.

Thus, while control theorists were first led to focus on the interpersonal milieu in which delinquent inclinations are formed, labeling theorists were first led to focus on the interpersonal milieu in which labels are imposed.

Earlier, less sophisticated versions of the labeling perspective seemed to suggest that the source of selective, discriminatory biases in the criminal justice system was the conscious, deliberate activity of the social control agents applying the sanctions. Although age, race, and class prejudice on the part of enforcers undoubtedly exists (and itself requires explanation), later research indicates that selective enforcement is also related to the specific nature of the offense, prior criminal record, and specific behavioral cues of the suspect as perceived by enforcement officials (Chiricos and Waldo 1975; Dannefer and Schutt 1982; Hewitt 1977; Piliavan and Briar 1964; Thompson and Zingraff 1981). These findings, of course, do not "refute" the labeling (or conflict or radical) perspective: (1) types of offenses are class related (e.g., property laws contain an inherent class bias protecting the propertied from the unpropertied); (2) prior criminal records can be plausibly considered to reflect previous selective enforcement bias; and (3) behavioral "cues" of criminality are no more intrinsically contained in suspects' behavior than is criminality in offender behavior. Instead of refuting the labeling perspective, such findings potentially reinforce it by offering specific areas of refinement.

However, a commonly acknowledged major shortcoming of labeling theory (see Gibbons 1979) is its failure to explain the social forms and patterns of individual behavior prior to label and sanction application. Many labeling theorists thought that this exclusion of consideration of "primary deviance" could be partially compensated for by their capacity to explain "secondary deviance," or that which occurred as a consequence of sanction application, stigmatization, changes in self-concept and self-esteem, and adoption of deviant roles (see Becker 1963; Douglas 1970; Goffman 1963; Lemert 1972; Lofland 1969; Matza 1969). Although incorporating a symbolic interactionist perspective (see Blumer 1969; Mead 1934) into explanations of deviant careers is certainly an advance (Davis 1975), the bulk of the studies conducted lend only scant and ambiguous support to the otherwise plausible secondary deviance thesis (see Foster, Dinitz, and Reckless 1972; Gibbons 1976; Mahoney 1974).

Labeling theory has made several essential contributions to the construction of a comprehensive criminological theory: (1) its recognition that no act is intrinsically deviant, (2) its renewal (see Mills 1943) of the rejection of presumed pathology in deviant behavior, (3) its incorporation of agencies of social control as a necessary focus of adequate criminological inquiry, and (4) its utilization of symbolic interactionist concepts linking individual behavior and immediate social milieus as reciprocal processes (see Becker 1963; Garfinkel 1967; Schur 1971). Nevertheless, the specific

areas of potential analytic refinement, the silence concerning "primary deviance," and the murky empirical evidence regarding "secondary deviance" imply that an adequate criminological theory cannot be confined to the labeling theorists' ahistorical, astructural preoccupation with the immediate sanction-application milieu and its role-reorganizing effects on the offender (see Davis 1972; Gouldner 1968; Liazos 1972).

Conflict theory and *radical criminology* arose as an aggregate of attempts to deal with some of the questions about the overall social structure left unanswered by labeling theory. These two schools are similar in their critical posture toward the prevailing social order but are distinguished by their conceptualization of the nature of the social order: (1) conflict theory tends to adopt a pluralistic, conflict perspective of society as an aggregate of diverse interest groups competing for power; and (2) radical criminology tends to adopt an instrumentalist-Marxian perspective of contemporary society as dominated by a unified, capitalist ruling class.

Following the leads of Sutherland (1947) and Sellin (1938), conflict theorists (Chambliss 1973; Hills 1971; Quinney 1970; Turk 1969; Vold 1958; Wolfgang 1968) reject conventional criminology's presumption that normative consensus characterizes contemporary society's formulation and enforcement of laws and contend that such legal controls arise in a pluralistic society of conflicting interest groups contending for power. The more powerful groups have the most influence in shaping the legal order to their interests, while those groups lacking social power are most frequently subjected to criminal status. Within the conflict perspective, criminality is preeminently a political question.

Rejecting the approaches of both contemporary criminology and the pluralist-conflict perspective, radical criminologists—including radicalized converts from the conflict perspective, such as Quinney (1974*a*, 1974*b*, 1975)—have attempted to develop an explicitly humanist-socialist critique of the capitalist social order (Platt 1974; Schwendinger and Schwendinger 1970; Taylor, Walton, and Young 1973). Viewing this social order, and the law formulation and enforcement arising out of it, as essentially dominated and controlled by a unified and cohesive (and implicitly evil) capitalist ruling class, such radical criminologists contend that the conventional narrow, legalistic conception of crime, espoused by both conventional and conflict criminological perspectives, is a liberal-technocratic, reformist conception and a form of tacit collusion with the immoral oppressive policies of the ruling class. Casting off such ideological constraints on criminological inquiry, radical criminologists postulate a counter-ideological reconceptualization of criminal behavior to include violations of fundamental human rights (crimes of exploitation and oppression) by the ruling class and its surrogates. Thus, not only is criminality to be seen as a preeminently political question; criminology itself is to be seen as an

521

intrinsically political endeavor (whether it be conservative, liberal, or radical).

Much of the substance of the radical critiques both of liberal-technocratic reformism and of the ideological distortions of the positivist-pathology model in conventional criminology we consider incisive and valid. Moreover, we agree that any comprehensive comparative assessment of contributions to "social harm" by the routine operations of capitalist institutions and upper-class individuals, and the aggregate "social harm" attributable to legally processed criminals, would weigh heavily against the former (Sutherland 1949).

However, given the patterns of victimization among the underclass and the ideological support for repressive measures promoted by street crime, radical criminologists can hardly celebrate street crime as a form of "primitive rebellion." In their passion for dispensing with conventional criminology's pathologized caricature of the criminal and its empiricist myopia regarding the larger social order in which criminality is situated, radical criminologists at first often simply substituted a romantic myopia regarding the criminal and a pathologized caricature of the social order. Later writings in radical criminology recognize this problem and begin to address the important ramifications that street crime has for buttressing class rule (Crime and Social Justice Collective 1976; Platt 1978).

Furthermore, in their discussions of both the politico-ideological distortions of conventional criminology and the class-biased operation of the criminal justice system, there often resides a simplistic suggestion of conspiracy on the part of the ruling class and its liberal-technocratic "servants" in the academic establishment. Invoking the specter of a ruling class conspiracy is hardly an adequate analysis of sources of power and authority in contemporary capitalist social formations, and it overlooks the distinction between manifest and latent functions of capitalist relations. These problems with radical criminology have been expounded on extensively from both Marxian and non-Marxian perspectives (Gibbons 1979; Greenberg 1976, 1981; Manning 1975; Rock 1974; Sparks 1980; Spitzer 1980; Steinart 1978; Sykes 1974; Turk 1975).

Despite their disparate pluralistic-monolithic caricatures of social power, the conflict and radical approaches both assume an untenable voluntarism with regard to whomever they view as powerful (authorities, interest groups, ruling classes, etc.). Paradoxically, it is as if the more powerful one is in exerting one's "will" to shape the social order, the more external one is to that social order—impervious to its structural constraints and omnipotent over it. Thus, if we are to explain the sources of social power, we have to examine the overall social structure in which it is differentially distributed. Neither conflict theorists nor radical criminologists have developed a coherent and refined conceptualization of social structure: the

former view it in the crude, ahistorical terms of eternal, arbitrary power disparities; the latter view it as an oppressive machinery manipulated by a ruling class.

Conflict theory and radical criminology would be considerably advanced by a fusion of certain dimensions of each perspective, accompanied by a mutual abandonment of their common voluntarist assumptions. Specifically, while conflict theorists should recognize the theoretical necessity to explicate systematically the sources of differential power and authority in their sociohistorically discrete forms and patterns, radical criminologists should recognize that power exercise and interest realization occur only in a process of struggle and are situated in a social structure characterized by multiple contradictions in which emerge political, economic, and ideological rivalries among distinct class fractions (Anderson 1974; Davis 1981; Marx 1963; Poulantzas 1978a). Commensurate with the abandonment of their voluntarist assumptions, both schools need to transport their penchant for interpersonal interpolations from the macro-level of structures to the more appropriate micro-level of genuinely interpersonal interactions. At this level, they would discover that many insightful contributions have already been made by conventional criminology. Significantly, shifts in precisely these theoretical directions can be discerned in more recent works of some Marxist criminologists (Chambliss 1975, 1979; Greenberg 1977, 1981; Quinney 1977; Schwendinger and Schwendinger 1976a, 1976b; Spitzer 1975).

Integrated theories that have recently emerged (see Elliot et al. 1979; Johnson 1979) attempt to tie together the theoretical fragments of conventional criminology. In addressing one of the theoretical silences of conventional control theory, integrated theory attempts to weave together the disconnected socialization settings of control theory into a unified conceptualization of socialization *processes*.

Retaining control theory's purely quantitative conception of bonding processes, integrated theories follow the individual through life-cycle encounters with various socializing agencies (families, schools, peer groups), inquiring at each step whether initially strong or weak bonds to the conventional order are being attenuated or reinforced as early socialization experiences "interact" with later ones. By incorporating additional concepts from strain, labeling, and learning approaches, integrated theorists attempt to reconstruct theoretically the individual processes of becoming, or not becoming, delinquent.

For instance, experiences of failure in school, and the attendant implications for labeling, can undermine already weak bonds developed within the family or create strain for juveniles with initially strong bonds, thus attenuating these bonds. Juveniles thus affected are more receptive to the influence of delinquent peer groups in which they learn specific

attitudes, motives, and skills to produce patterned delinquent behavior and possibly consolidate compensatory delinquent identities through alternative bonding processes.

Integrated theories do not view the delinquent outcomes of socialization sequences in undifferentiated terms. Rather, integrated theories allow us to develop qualitative and quantitative distinctions for delinquency (e.g., instrumental, violent, occasional, or patterned). These distinctions are made theoretically possible by the incorporation of concepts from various criminological theories, each reflecting different dimensions of the development of delinquent behavior; the qualitative and quantitative variations in delinquency are made etiologically possible by the multiple paths of socialization through which a juvenile may pass.

Needless to say, integrated theories are a major advance of micro-level inquiry into the interpersonal interaction processes producing delinquencies. Nevertheless, precisely because integrated theories have thus far been restricted to combining the prevailing theoretical currents in conventional criminology, they have also incorporated some of the latter's oversights and distortions. Thus, integrated theories uncritically incorporate the silence of control theory on the social structural origin of initial bonds produced in families. Bonds are conceptualized in the quantitative terms of strong or weak bonding to a conventional order. Moreover, and related to these oversights, integrated theories fail to transcend the confines of micro-level, interpersonal processes to examine why the various socialization paths are arrayed as they are and why distinctive socialization sequences are differentially distributed along class lines.

Such questions cannot be addressed within the confines of micro-level theory—even an integrated theory incorporating a series of micro-level, interpersonal processes. In order to address the systematic, processual character of interpersonal interactions—the larger mosaic of micro-level patterns—we must postulate subapparent but encompassing structures of social activity. Just as the *path* of electrons cannot be entirely explained by their apparent motions but only by postulation of a subapparent structural (electromagnetic) field determining a *process* of particle movement (Einstein and Infeld 1938)—so, too, the path of socialization through which individuals pass cannot be entirely explained by their apparent interpersonal activities. Such explanation requires the postulation of a structural field. This social structural field is postulated to contour the relations of social control, pattern the arrangement of rewards and punishments, and thus determine the distinctive paths of socialization available to individuals. Insofar as this social structural field determines the forms and contents of socialization experiences specific to families, schools, and peer groups, it determines the patterned processes of development of both delinquent and nondelinquent behavior. A postulated social struc-

tural field of class relations allows us to develop precise distinctions in socialization paths that are not randomly distributed but are connected to the underlying relations of commodity production in capitalism. It is to such a field theory, incorporating structural-Marxist concepts, that we now turn.

AN INTEGRATED STRUCTURAL-MARXIST APPROACH

The processes of socialization described by integrated theorists can be placed within a context of larger structural relations that define these processes. Structural-Marxists[4] have attempted to connect relations of production theoretically with those relations necessary to reproduce a social formation. It is within this theoretical framework that the latent social production of delinquency is best understood.

In a Marxist paradigm, the fundamental structural relations are those that are entered into at the point of material production. All other human relationships rest on those relationships that revolve around the provision of the physical means of life. Thus Marxists consider the mode of production and reproduction to be the essential starting point of any social analysis, because human existence depends first on the biological survival of the species and thus on the meeting of human needs for which material resources must be expropriated from nature. The relationships among human actors that arise from these essential productive tasks set limits on other, nonproductive social relations, without necessarily determining in any mechanistic sense their form and content.[5] Entering into these necessary productive relationships not only produces the means for physical subsistence but also produces and reproduces the practical intercourse between human beings, that is, produces and reproduces the structured relations through which human actors interact in the essential tasks of material production (Marx and Engels 1967, pp. 6–7).

Individuals do not enter into these relations arbitrarily or simply as a product of their will but encounter them as a given product of the historic development of productive forces. Individuals must interact with these relations (and change them) within the structural limits set by the relations. As Marx and Engels (1967, p. 18) state, "The social structure and

---

[4] Early formulations of structural-Marxism include those of Althusser (1971, 1977) and Poulantzas (1978b), all of which suffered from ahistorical, static conceptions of structure and failed to include a dialectical notion of the relation between structure and human agency as mediated by class struggle (Anderson 1980; Appelbaum 1979; Burris 1979; Laclau 1979; Wright 1978). A provocative critique of structural-Marxism is provided by Thompson (1978). Anderson's (1980) reply to Thompson provides an incisive restatement and the most sophisticated elaboration to date of the structural-Marxist position.

[5] For a concise statement of a Marxist understanding of structural determination, see Wright (1978, chap. 1).

the State are continually evolving out of the life-process of definite individuals, but of individuals, not as they may appear in their own or other people's imagination, but as they really are; i.e., as they are effective, produce materially, and are active under definite material limits, presuppositions and conditions independent of their will."[6]

What most clearly distinguishes structural Marxism from the more instrumentalist brands of Marxism is the degree to which actions by ruling (or any other) class members are considered to be voluntary and in concert with other members of the same class. For instance, under capitalist relations of production, even capitalists, as private owners of the means of production, are in a relation of competition with each other which determines the limits on their actions and often necessitates that certain measures be instituted. Even monopolistic practices, such as price fixing, are necessitated by the vicissitudes of capitalist competition. Thus the relations of private ownership of the means of production set certain constraints on capitalists as they interact with each other and as they interact with workers from whom they expropriate surplus labor.[7]

These relations of production also define the class structure at any point in history. Structural Marxism conceptualizes classes in social structural rather than interpersonal terms: classes are integral, not external, to social structure; they are defined by the social relations of production, specifically by relations to the means of production. Classes can be distinguished by their ownership and control of the means of production and of its end products. This differs from defining classes from the point of view of attributes of individuals, for example, life chances, education, income, or occupational prestige, though these may be related to the structural definition of class.[8] As such, classes exist in relation to each other only vis-

[6] This is not an ahistorical abstraction or law of human nature. As Marx (1981, pp. 957–59) discusses and Anderson (1980, chap. 2) reiterates, the still-persisting and most fundamental structural constraint on human action is material scarcity. As the forces of production are expanded to a point where material scarcity becomes less of a constraint, human agency takes on a more determining role.

[7] Greenberg (1981, p. 14) defines "surplus labor" as "labor [or its products] above and beyond what is needed to 'reproduce' the laborer from day to day and from generation to generation . . . the form in which it is appropriated varies greatly [from society to society]." Greenberg (1981, p. 28) goes on to explain: "In the capitalist mode of production, workers sell their labor-power to capitalists in return for wages. The difference between the value that workers receive in the form of wages and the value they produce is surplus value—the specific form that surplus labor takes in the capitalist mode of production—and is appropriated by the capitalist. This is the source of profits." As capitalists compete to expropriate higher levels of surplus value from workers, wealth, used to produce further wealth, increases in a process known as capital accumulation. This process gives the capitalist mode of production its dynamic impetus.

[8] See Wright's discussion (1979, pp. 3–18) comparing the structural-Marxist conception of class, focusing on production relations, with other conceptions of class. Employing such a Marxist, "relational" conception of class—in which most "white-collar" workers are clas-

à-vis the means of production. And as such, the activities of class members, of ruling or subordinate classes or class fractions, are determined by (1) the structure of the mode of production and the patterns of its development; and (2) the relations of that class or fraction with other classes or fractions, which are often antagonistic relations because of structural contradictions characterizing the mode of production.

The modern class structure is a product of a continuous process of struggle among classes in the historical development of capitalism. Figure 1 presents the general configuration of class relations in modern U.S. capitalism. An extended discussion of this class structure is necessary in order to clarify the role of the larger social structural context in the production of serious, patterned delinquency.

The modern class structure arose through a process of struggle among three major classes: the capitalist class, the working class, and the petty bourgeoisie. These classes are placed in relations of antagonism due to the underlying imperatives of capital accumulation. These antagonisms are mitigated or intensified by specific economic, political, and ideological control structures. The antagonistic relations that arise from exploitation in the accumulation process may be conceptualized (metaphorically) as a centrifugal force that pushes the social formation toward disintegration. This underlying class antagonism periodically erupts in open struggle; such eruptions necessitate the establishment (if historical conditions permit) of specific control structures to temper the antagonism. The control structures, when effective, obscure the underlying class struggle. They can be seen (again metaphorically) as a centripedal force that pulls the social formation toward integration. These control structures can be, and have been, upset by periodic economic crises which undermine the efficacy of the controls (such as severe recessions which disrupt wage concessions that ameliorated open struggle between capital and labor and which intensify competitive relations between sectors of capital, disrupting certain political accommodations). The contradiction and constant interaction between these disintegrative and integrative forces define the historical movement of capitalism and, more specifically, determine alterations in the class structure.

Without the connotations of normative consensus invoked by the term "society," modern society can be conceptualized as a *social formation* (see

---

sified as part of the proletariat or working class, rather than as part of the "middle class" as in most SES measures—Wright is able empirically to predict differences in income better than do those models based on conventional SES criteria. Also see Braverman (1974), Edwards (1979), Levison (1974), and Parker (1972) for discussions and critiques of conventional SES measures of class.
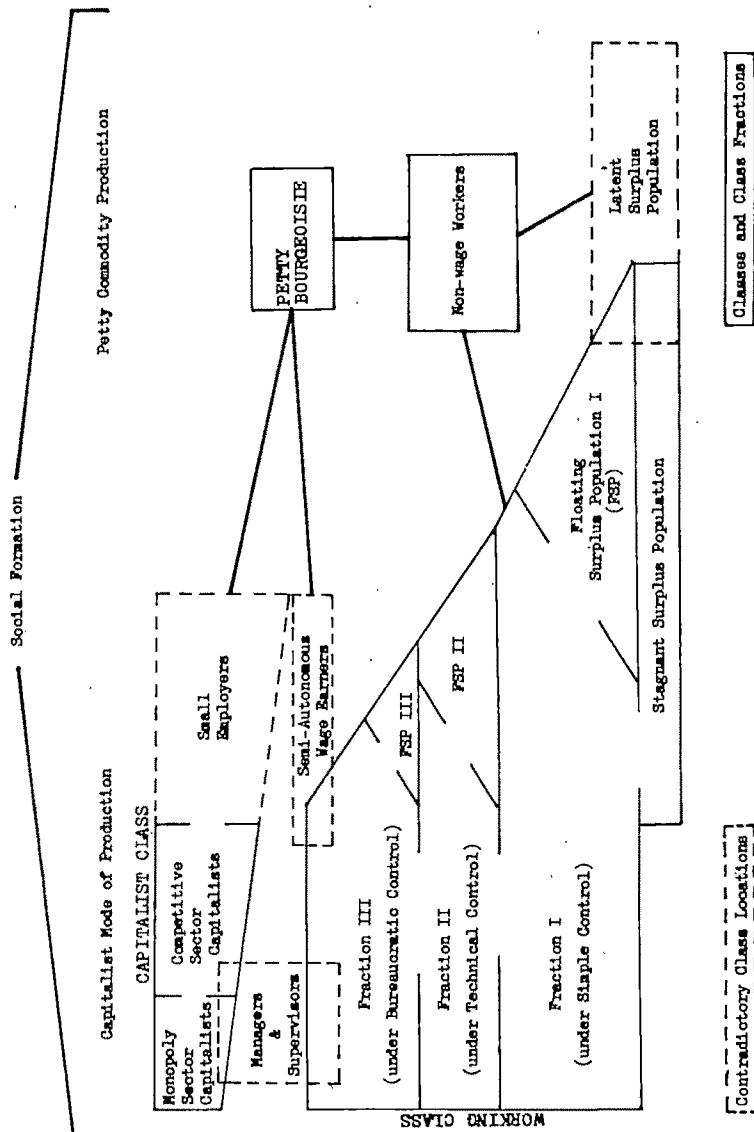
FIG. 1.—The class structure of the United States (adapted from Braverman [1974]; Edwards [1979]; Wright [1978]; material used with permission of Monthly Review Press, Richard Edwards, and New Left Books).

Greenberg 1981; Poulantzas 1980) composed of two modes of production: capitalist and petty commodity production. In petty commodity production, the direct producer obtains the means of subsistence by selling on the market commodities that he or she has independently produced with the possible help of family members, nonwage apprentices or tenant workers, and only a few, if any, wage laborers. In contrast, the capitalist mode of production is characterized by (partially socialized and interdependent) production of commodities by direct producers who are separated from ownership and control over the means and ends of production, and who must sell their labor power for wages in order to obtain the means of subsistence. Limited by its less productive technical means and small-scale division of labor, petty commodity production, in which the roles of owner of the means of production and owner of the final product are fused in the family unit with the roles of laborer and consumer, is geared toward smaller-scaled, simple reproduction of the family unit. However, market competition between producers engenders an expansionary thrust in which direct producers are compelled to introduce innovations in technology and expand the division of labor, or they are increasingly separated from the means and ends of production; thus the fused roles in the petty producer family are fractured into the owners, managers, workers, and consumers characteristic of capitalism (Post 1982). It is this separation of roles and commensurately intensified competition among those in each role that provide capitalism with its incessant expansionary dynamic (and with its recurrent overproduction crises). Thus, petty commodity producers predominate in areas where profits are low or precarious (such as farming), and competition subjects petty commodity producers to perpetual incipient proletarianization and is the underlying source of struggle between the capitalist class and the petty bourgeoisie (Marx 1977).

The intrusion of the capitalist mode of production on petty commodity production (as with the rapid capitalist expansion into agriculture after World War II) involves the production of a latent surplus population (Braverman 1974), composed primarily of agricultural laborers, tenant farmers, and family farm workers, who are displaced by technical innovations and corporate intrusions into agriculture and other types of petty commodity production. The latent surplus population finds itself midway between petty commodity production (from which it has been pushed) and the capitalist mode of production (toward which it is pulled for survival). It may be seen as a "contradictory class location."

Wright (1978) developed the notion of "contradictory class locations," which we apply to the latent surplus population, in connection with two other locations that stand between the capitalist mode of production and petty commodity production. They are contradictory because they share opposite objective interests of two antagonistic classes. They are captured

529

within the struggle between classes and are defined by their contradictory location in this struggle. First, small employers enjoy the autonomy of the petty bourgeoisie that comes with the ownership and control over the small-scale means of production and its end product. At times, because of their smaller scale of production, small employers, like the petty bourgeoisie, are threatened by competition and intrusions from big capital. Yet, just as the capitalist class, they employ wage labor and are engaged in struggles over workplace control and labor costs with segments of the working class. Second, semiautonomous wage earners, such as college professors and university research scientists, share with the petty bourgeoisie greater control over their own labor process and over disposal of their final product. Yet, just as the working class, they are in a wage-labor relationship with their employer that at times leads to organized struggle against that employer.

Since petty commodity production continually gives way to capitalist incursions, the expansionary dynamic of capitalism renders the latter the dominant mode of production in the modern United States. Yet, even the capitalist mode of production has not generated two monolithic classes at this point in history. Both the capitalist class and the working class are composed of class fractions that have evolved with the development of capitalism.

Fractions of the capitalist class have emerged through the struggle between capitalists for market shares, which has resulted in an uneven expansion of capital and the corresponding rise of sectors within the capitalist economy (Edwards 1979; O'Connor 1973; Oster 1979; Tolbert, Horan, and Beck 1980). Some capitalists have been able to capture greater market shares and monopolize certain industries, such as mining, transportation, utilities, communications, and some manufacturing industries (e.g., auto, steel, electronics, and petroleum). The monopoly-sector capitalists are better able to set noncompetitive prices in concert with other monopoly capitalists and to accumulate capital on a larger scale (Baran and Sweezy 1966). Thus they have greater (but not exclusive) power to influence state policies and are in a better position to temper working class resistance through wage concessions and more sophisticated methods of workplace control. Foreign competition (as in the auto and steel industries) presents the greatest structural constraint on the monopoly sector.

It is primarily in the monopoly sector of capital that top managerial and supervisory positions have been established to provide the more sophisticated control networks within organizations and to deal with complicated market arrangements that may have international scope. These new managerial and supervisory positions have been identified by Wright (1978) as a "contradictory class location" between the capitalist class and

the working class. On the one hand, persons in this location share with capitalists the control over the means of production and control of the sale of end products. On the other hand, they lack the ultimate control over decisions that goes with direct ownership. Managerial and supervisory control is contingent on the profits that accrue to the owners as a result of these decisions. When profits drop as the result of top management's miscalculations, it is the owners' prerogative to dismiss the managers from employment, just as the owners would dismiss workers (Bauer 1981).

The other major fraction of the capitalist class is composed of competitive-sector capitalists. These capitalists are less protected than the monopoly sector from market fluctuations; they are concentrated in wholesale, retail, and service industries and in a few manufacturing industries, such as lumber, textiles, and apparel. They cannot set noncompetitive prices owing to their relatively greater vulnerability to market forces, and accordingly are not able to accumulate capital on as large a scale as the monopoly sector. Thus competitive-sector capitalists lack the resources for wage concessions and generally cannot afford the more sophisticated managerial techniques (and personnel) to control their labor force. Instead, they must depend more heavily on the coercive control provided by the market force of supply and demand for labor to keep wages down and to use threats of dismissal effectively as the primary measure for disciplining workers. The floating surplus population (Braverman 1974), those irregularly employed workers who are readily called up as workforce replacements, provides the competitive capitalist sector with its basic tool for control over its work force. The stagnant surplus population (Braverman 1974) is less actively involved in competition for jobs and moves between unskilled, extremely low-paying jobs, usually in service industries, and state controls provided by welfare and criminal justice apparatuses (Colvin 1981; Piven and Cloward 1971; Quinney 1977).

The different market positions of the monopoly and competitive sectors of capital present different structural limitations on each sector in its ability to resist, meet, or anticipate the demands of their workers. Within the context of these structural limitations, capitalists institute specific structures of control over the labor process in response to the level of struggle presented by workers. Market competition compels capitalists to introduce technical innovations and more efficient divisions of labor (Braverman 1974). But these innovations often meet with resistance from workers who struggle for control over the production process (Montgomery 1979). The various methods used to control workers, which are historically developed through the struggle with workers, on the one hand, and through imperatives of market competition, on the other, define the major fractions of the modern working class.

531

Edwards (1979) presents the most comprehensive analysis to date of the major fractions of the working class. Fraction I is composed of workers who labor in more competitive industries and who, in Edwards's terminology, are involved in a "secondary labor market," which operates largely by the law of supply and demand for labor in setting the price of labor power. Thus, fraction I is most affected by, and includes a huge proportion of, the floating and stagnant surplus populations which compete for fraction I jobs. These include low-skill jobs in small, nonunion manufacturing; southern textile jobs; service jobs; lower-level clerical and sales jobs; and agricultural wage-labor. These jobs are "typically dead-end jobs, with few prospects for advancement and little reward for seniority in the form of either higher pay or a better job" (Edwards 1979, p. 167). Employers forcefully resist organizing efforts by workers. There is a large turnover of employees in fraction I, which undermines labor-organizing efforts and also reflects the predominant control mechanism for worker discipline—threatened or actual dismissal from the job.

Edwards defines the control mechanism used to discipline fraction I workers as "simple control," which, in our view, corresponds to Etzioni's (1970) "coercive compliance structure." This compliance structure rests predominantly on "coercive power," which Etzioni (1970, p. 104) defines as involving "the application or the threat of application of physical sanctions" that include "the controlling through force the satisfaction of needs." Dismissal from the job definitely involves the forced removal from the means for satisfying basic needs and thus is a coercive mechanism of power. Coercive power tends to create an "alienative involvement," or intense negative orientation (Etzioni 1970, p. 106) on the part of the worker toward the employer and the organization. This workplace control structure conditions the worker's ideological orientation or bond to authority at the workplace. In this case an alienated or negative bond is created by the more coercive simple control used with fraction I workers.

Fraction II is composed of organized workers who, through earlier struggles, gained wage and benefit concessions, job protection, and the establishment of industry-wide unions. The auto, steel, and rubber industries; machine manufacture; and mining offer the best examples of fraction II workers. According to Edwards, capitalists and their managers were forced to shift from simple control to more sophisticated technical control in response to organized worker resistance to the arbitrary and coercive discipline of simple control. Technical control is machine paced and impersonal and relies on the worker calculating his or her material self-interest for pay raises and job security based on seniority with the firm. Edwards identifies fraction II workers as being involved in a "subordinate primary labor market" that relies very little on external pressure from a floating surplus population, which is usually protected in this

fraction by more generous, union-connected unemployment benefits. Job performance is usually routine and boring and holds little intrinsic interest for the worker. The main control is the calculation of extrinsic rewards of material security through movement up a union pay ladder; this process produces a firm-specific "internal labor market" (Edwards 1979, p. 172). This form of control, which defines fraction II, is consonant with Etzioni's (1970, p. 109) "utilitarian compliance structure" composed of "remunerative power," involving the manipulation of material rewards. A "calculative involvement" of intermediate intensity on the part of the worker is produced in this type of control structure. It is a precarious ideological bond, depending on continual remuneration and advancement up the pay ladder and producing little loyalty on the part of the worker.

Fraction III is defined by its involvement in an "independent primary labor market" composed of jobs that are likely "to require independent initiative or self-pacing" (Edwards 1979, p. 174), including: (1) middle-layer workers, such as technical staff, foremen, personal secretaries, and middle-level supervisors; (2) wage-earning craft workers, such as electricians, carpenters, plumbers, and machinists, who are usually members of craft guilds; and (3) salaried professional workers, such as corporation accountants, corporation research scientists and engineers, corporate and state lawyers, social workers, and school teachers. The bulk of state workers are included in fraction III. This fraction, just as other working-class fractions, spans both blue-collar and white-collar occupations. The jobs in fraction III are ruled by more universal standards of professional or craft guild conduct. This differs from the machine-paced and firm-specific standards of the "subordinate primary labor market" of fraction II.

The floating surplus population plays a very small role in controlling fraction III workers, since higher professional and educational requirements prevent most workers from competing for these positions. Edwards (1979, p. 131) argues that, after World War II, many new monopoly industries (computers, technical instruments, chemicals, refining) adopted many of the aspects of technical control and added other sophisticated control mechanisms to ward off the possibility of industry-wide unionization. In any event, the work tasks of most fraction III workers preclude the use of pure technical control, since the work tasks cannot be easily routinized, requiring instead a certain degree of flexible response and initiative on the part of the worker.

"Bureaucratic control" (Edwards 1979, p. 131) is thus necessary to gain compliance from fraction III workers, whose products, as with any wage laborer's product, are expropriated for the profit of the owners. This lack of control over the end product of one's labor is what distinguishes fraction III from semiautonomous wage earners. Bureaucratic control involves an

elaborate manipulation of symbols and statuses that elicits ideological commitments to the organization. The bureaucratic control structure "fosters occupational consciousness; that is, [it] provides the basis for jobholders to define their own identities in terms of their particular occupation" (Edwards 1979, p. 177). Thus, fraction III jobs hold more intrinsic meaning for these workers. While pay scales are commensurate with increases in job status, it is primarily the possibility of job status increase that compels worker compliance. In fact, many fraction III workers are paid less than unionized fraction II workers.

"Bureaucratic control" corresponds to Etzioni's "normative compliance structure" that includes "normative power," which "rests on the allocation and manipulation of symbolic rewards" (Etzioni 1970, p. 104) and elicits a "moral involvement," designating "a positive orientation of high intensity" on the part of the worker toward authority and the organization (Etzioni 1970, p. 107). While ultimate decisions about the organization are left to owners and top management, status among fraction III workers is based on the degree of autonomy and decision-making power allowed for each position. Status comparisons with peers provide a strong element of control. Thus many of the job positions in fraction III overlap with the contradictory class locations of top managers and supervisors and of semiautonomous wage earners. (A good example of the latter overlap is Bell Laboratory scientists who are allowed to spend 10% of their time working on their own inventions.) The job experience of fraction III workers, in which more complex tasks are encountered and greater autonomy and decision making are exercised, conditions these workers' ideological orientation toward a more positive bond to their work, to authorities at the workplace, and to their organization.

The three fractions of the working class are based on the experiences that arise from the three primary structures of workplace control. It is the objective structures of control that condition subjective class experiences and give rise to class-related ideological orientations. These objective control structures, which arose through the historic development of capitalist relations of production, are the crucial link between the mode of production and subjective class experiences. It is the structure of workplace control that affects the ideological orientation (or bond) of the worker, who in turn contributes to the production and reproduction of a specific type of control structure in the family, the initial institution in which the reproduction of class relations and labor power takes place.

Family Control Structures and Delinquency

Melvin Kohn (1977) argues that a parent who experiences greater external control at work comes to "understand" the importance (for physical and

financial survival) of conformity to external authority. The stress on conformity to external authority is then (consciously or unconsciously) impressed on children during child-rearing activities in the family. In contrast, parents who experience a lower degree of external control and exercise greater decision-making power at work come to view "internalized self-control," initiative, and creativity as valued attributes that are impressed on their children.

Both orientations for parental values spring from the parents' ideological bonds produced in differential work experiences. An underlying message to the child in both instances is an ideological statement about the world: control of life circumstances and the determinants of one's behavior spring from either external compulsion or internal motivation. The child, in his or her everyday interactions with parents, learns that one acts toward authority either out of fear or calculation of external consequences or out of a sense of internalized respect or commitment. Through the process of parental control over children, a parent's bond to the authority of the workplace is reproduced in the child's initial bond to parental authority.

Thus Kohn (1977) found in his national survey that lower-SES parents and parents under more tightly controlled and routinized work situations tend to have more alienated ideological bonds, impress the attributes of conformity to external authority on their children, use more punitive and physical disciplining practices, and punish their children more for consequences of their acts than for perceived intent of their behavior. Higher-SES parents and parents from more "self-directing" work situations show opposite directions in associations to the bonding, parental value, and parental disciplining variables.[9]

The child participates in a family control structure that contains certain rewards and punishments for specific types of behavior and perceived motives. The more coercive this structure of control, the more negative are the bonds produced in the child. The degree of coerciveness of family control structures is influenced by material resources available to parents as reinforcements and by the ideological bond of the parent, which influences the types of child behavior that will be rewarded or punished. Both material resources and parental bonds are associated with workplace control structures.

As with the parents' social control experiences at work, an initial ideo-

[9] While statistical associations are significant, many are statistically weak (Erlanger 1974). The reader should, therefore, see these findings as suggestive only. Kohn's findings have been replicated by several quantitative studies (Coburn and Edwards 1976; Erlanger 1974; Franklin and Scott 1970; Gecas 1979; Gecas and Nye 1974; Kohn 1976; Kohn and Schooler 1973; Pearlin 1970; Wright and Wright 1976) and by two important qualitative studies on family and work (Komarovsky 1962; Rubin 1976).

logical orientation either toward conformity to external authority, based on fear or calculation of external consequences, or toward internalized "self-direction" is produced in the child, depending on the coerciveness of the control structure within the family. This family control structure is shaped by the stability and level of parents' association with control structures at work.

Thus we postulate that parents from fraction I of the working class, typified by more coercive workplace controls and more sporadic associations with specific workplaces, tend to enforce an uneven and erratic family control structure that swings unpredictably between being lax and being highly punitive. (See Rubin's [1976] discussion of "settled living" versus "hard living" life-styles within the working class.) We expect more alienated initial bonds to be produced in children who experience such arbitrary, inconsistent, and coercive family control structures. Parents from fraction II of the working class, characterized by more steady and long-term association with a utilitarian compliance structure at work, tend to enforce a more utilitarian compliance structure in the family that produces calculative bonds of intermediate intensity in their children, who can dependably predict external consequences for behavior. Parents in more "self-directing" workplace situations—such as fraction III workers, semiautonomous wage earners, top managers, and owners of the means of production—tend to enforce a more normative family compliance structure in which positive initial bonds of high intensity are produced in children.

Though no studies have been designed that directly link workplace controls, family disciplining practices, and delinquency, some empirical data suggest that our postulations may be correct. In addition to Kohn's (1977) findings, Strauss, Gelles, and Steinmetz (1980) in a national survey of self-reported family violence found significant inverse associations between class-related variables (income and occupational prestige) and parental violence toward children and a significant positive association between unemployment and parental violence toward children. These findings suggest class differences in coerciveness of family control structures.

Regarding delinquency, significant positive associations between more physical and punitive parental disciplining practices and delinquency and between erratic parental disciplining practices and delinquency have been reported in studies using both official and self-reported delinquency measures (Glueck and Glueck 1950; Hirschi 1969; McCord, McCord, and Zola 1959; Nye 1958; West and Farrington 1973). Hirschi (1969) reports that juveniles' positive bonds to parents are inversely related to self-reported delinquency; later studies (see Johnson 1979) suggest, however, that the influence of parental bonding on delinquency is at best indirect.

These studies suggest that the coerciveness of family control structures, conditioned by parents' work experiences, contributes at least indirectly to the production of delinquency.

School Control Structures and Delinquency

The next important process in the generational reproduction of labor power, necessary for capitalist production relations, is formal schooling. Education entails the molding of human raw material into a product that can be consumed in the capitalist labor process. This product must not only possess certain technical skills but also be appropriately disciplined and ideologically amenable for entry into the various compliance structures at work. Educational institutions serve these purposes; control structures at schools have been designed to fit the labor requirements of capitalist enterprises (Bowles and Gintis 1976; Carnoy and Levin 1976; Noble 1977; Spring 1972).

When entering school, the child, with initial bonds produced in a family control structure, confronts a new structure of control. The school, like parents' workplaces, contains gradations of control (within various "tracks") that are exercised over students. A child with negative initial bonds is likely to be placed in a control structure at school that parallels the coercive family control structure that produced the child's negative bond. This process operates both by deliberate design and through subtle mechanisms.

First, IQ and aptitude tests may designate negatively bonded children for processing in "lower-level" tracks that are more regimented and coercive. The aptitude tests may measure initial bonds more than they measure innate intelligence, for the IQ test "probably measures a rather special kind of motivation" (Jencks et al. 1972, p. 55). Since there is no extrinsic reward or punishment associated with giving right or wrong answers on these tests (nor any feedback on how well one has done), those students who do not give random answers must be internally motivated to achieve. Thus a high IQ score denotes a positive internalized bond that was produced within a family control structure. The IQ score is a primary mechanism for placement into educational tracks which contain differential structures of control.

A second mechanism is the reciprocal relationship between a child's initial bond and the child's elicitation of controls from teachers and other authority figures in the school setting (as with Piliavin and Briar's [1964] finding of police reaction to behavioral cues of juveniles). The negatively bonded child, for instance, may give behavioral cues of being (in the teacher's perception) potentially disruptive. These behavioral cues may elicit a labeling process that creates the "self-fulfilling prophecy" of a

537

disruptive child who elicits still more coercive controls than before. The labeling process can be seen as a reciprocal relationship between differentially bonded individuals and agents of control structures. The process of labeling in educational settings has been demonstrated experimentally (Leacock 1969; Rist 1970; Rosenthal 1973; Rosenthal and Jacobson 1968; Rubovits and Maehr 1971), though other researchers have been unable to replicate these findings (Barber et al. 1969; Gephart 1970).

A third mechanism involves the finite number of positions in various educational tracks that may not correspond with the number of juveniles who are motivationally prepared (owing to initial bonding) for more "self-directing" tracks. This situation would present a more immediate structural source of strain at all tracking levels for those children who are placed lower than initial family or early school socialization led them to expect. As in Elliot et al.'s (1979) integrated theory, these juveniles, because of the structural disjuncture, would experience failure and strain that would attenuate positive bonds and lead to possible placement in more coercive school control structures.[10]

A fourth mechanism is the differential financial resources of schools in lower-class, working-class, and middle-class neighborhoods. This difference in resource allocation springs largely from differential tax bases associated with income and property ownership differences. This circumstance creates differences in the availability of rewards and punishments for controlling children in the school setting and often necessitates a greater reliance on coercive measures in lower-class schools. Thus, tracking takes place between, as well as within, schools.

The various tracking experiences in the school correspond to the social control requirements at workplaces (Bowles and Gintis 1976). The lower tracks emphasize strict discipline, regimentation, and conformity to external authority. The higher tracks emphasize initiative, creativity, and self-direction. Again, these control structures with their various rewards and punishments shape experiences for individual children and reinforce or attenuate initial bonds to authority.

Thus, parents' bonds engendered by workplace control structures are reproduced in the child through family controls; the child's initial ideological orientation toward authority is most likely to be reinforced in school control structures by mechanisms which render the child suitable for eventual placement in a workplace control structure similar to that of the parent. Thus, inequality and the class structure are subtly reproduced.

Some research suggests that school control structures and related vari-

[10] This situation would constitute in Etzioni's (1970) terminology an "incongruent compliance structure" that would tend toward a "congruent" coercive compliance structure by producing an alienative involvement on the part of the juvenile.

ables are associated with delinquency. Hirschi and Hindelang (1977) report that one of the most consistent relationships found in delinquency literature is the inverse association between IQ scores and delinquency. Also, positive bonding to school has been associated with lower levels of delinquency (Hindelang 1973; Hirschi 1969; Johnson 1979). In addition, lower educational track position has been associated with higher rates of delinquency (Kelly 1974; Schaefer, Olexa, and Polk 1972). However, Wiatrowski et al. (1982), using a sample of high school students, found a significant but very weak inverse relationship between curriculum track and delinquency. It may be, as Wiatrowski et al. (1982, p. 158) point out, that tracking experiences have their impact on delinquency earlier than high school. Indeed, as Greenberg (1977) notes, involvement in serious property offenses (which constitute the bulk of the FBI's Part One Index crimes) peaks around ages 15–16 and then declines, suggesting that the increased regimentation and sudden intensification of tracking that accompanies the shift from elementary to junior high school may be more crucial in explaining delinquency than are later tracking experiences in high school. For explanations of sustained involvement in serious delinquency after junior high school, we must examine the growing influence of peer associations.

Peer Associations, Structures of Control, and Delinquency

Peer associations mediate the relationship between delinquency and the other control structures we have discussed. The structured experiences associated with parents' workplaces, families, and schools form the necessary conditions and delinquent peer associations create the sufficient conditions for patterned delinquent behavior.

Peer groups form largely around juveniles' interactions at and around school and reflect the patterns of association structured by school tracking systems and by the differential distribution of opportunities among neighborhoods. Structures of control over adolescents, who are marginal to workplace control structures and who are increasingly moving away from the influence of family control structures, are shaped by a combination of school, neighborhood, and peer group relationships (Greenberg 1977).

Our theory assumes that juveniles with similar bonds (due to similar experiences in schools, neighborhoods, and families) will be in greater contact with each other and will be drawn to each other on the basis of shared values and experiences. Specifically, students in higher educational tracks, in which positive bonds are reinforced, tend to form peer associations that contain more normative compliance structures that continue the production of positive bonds and thus insulate these juveniles from delinquent involvement. Students in intermediate tracks, which are more

539

conducive to calculative involvement, are likely to form peer relations oriented around the pursuit of extrinsic rewards. The possible contradiction between rewards in school and those associated with peer relationships places these juveniles under greater strain. Such a situation can attenuate their bonds and lead to occasional involvement in less serious types of instrumental delinquency. Students in lower educational tracks, which tend to produce more alienated bonds, would be most open to associating with similarly negatively bonded peers who are likely to reinforce tendencies toward serious, patterned delinquency (Elliot et al. 1979).

Studies strongly suggest that most delinquent behavior occurs in groups and that juveniles with delinquent peer associations tend to be delinquent themselves (Erickson and Jensen 1977; Short, Rivera, and Tennyson 1965; Short 1958; Voss 1964). There is conflicting evidence on the relationship between bonding among peers and delinquent behavior. Hirschi (1969) reports that "attachment" (or positive bonding) to peers is inversely associated with delinquency. Jensen and Erickson (as reported in Jensen and Rojek 1980) found no association between attachment to peers and delinquency. Yet Hindelang (1973), Empey and Lubeck (1971), and Erickson and Empey (1965) found positive associations between attachment to peers and delinquency. These disparate findings may be due to different measures of "attachment to peers" (Jensen and Rojek 1980) and to differences between rural and urban samples in some of these studies. We argue that a more important source of disparity in these findings may exist: attachment between delinquent peers may vary by the seriousness and frequency of delinquent acts in which a peer group is involved. We postulate that control relations within delinquent peer groups determine both the level of bonding between the peers in question and the type of delinquency in which these juveniles are involved.

This latter point is suggested in Cloward and Ohlin's (1960) description of types of delinquent "subcultures," two of which we will discuss. The "criminal" delinquent subculture is oriented toward instrumental delinquent activities, such as theft. We argue that the emphasis placed on material gain by members of this delinquent subculture emanates from a utilitarian control structure contained within the instrumental delinquent group itself. Maintaining access to material rewards becomes important for establishing a position within the internal control structure of this type of delinquent group. Group members with stronger "connections" to adult fencing operations or to adult professional thieves, and group members who are able to "haul" more goods, enjoy greater status and power within the group because of their access to sources for remunerative rewards. Thus internal competition for positions of control within the group requires illegal activities oriented around material gain. This internal competition meets with an external social environment in

which legitimate instrumental opportunities are constrained, but illegitimate instrumental opportunities are open for those "cunning" enough to exploit them. These internal and external forces combine to produce the structural inducements for serious, patterned, instrumental delinquency characteristic of the "criminal" subculture. The internal control relations within these instrumental delinquent groups are conducive to a type of bonding based on calculation of material rewards and punishments; thus an intermediately intense, less negative bond may emerge among these juveniles, causing their mutual attachments to be stronger than those of other delinquent peers.

The second delinquent "subculture" described by Cloward and Ohlin (1960) is labeled the "conflict" subculture. This specific type of delinquent peer association, according to Cloward and Ohlin, arises from the absence in the immediate social environment of both legitimate and illegitimate opportunities for material gain. We would expect this type of delinquent group to be predominant in urban areas where the surplus population has been most recently pushed from agrarian social and economic relations and is least established as a genuine urban community.[11] The delinquent activities of these "conflict" groups are structured around violent gang rivalries over dominance within neighborhoods, often in nascent attempts to establish and control an illegitimate opportunity structure. The lack of material resources for control, and the associated lack of legitimate opportunities for status achievement, render control relations among peers within these groups more coercive. Thus, bonds among peer group members are negative, while, at the same time, violent behavior and motives are reinforced in interactions within and among "conflict" delinquent groups. Those members who gain reputations for being the most violent are able to exert the most control over the group. Again, we see an interaction between internal relations of group control and external structures within the surrounding social environment, in this case an interaction between internal coercive control and external competitive violence between groups. These internal and external forces combine to create the structural inducements for serious, patterned, violent delinquent behavior. These conflict delinquent groups resemble Yablonsky's (1962, 1966) "near groups," characterized by coercive, authoritarian leaders and an absence of solidarity among group members who are negatively bonded to each other.

[11] Because ethnic criteria provide the most readily convenient, invidious distinctions for designating constituents of the surplus population, such urban areas are usually populated by diverse ethnic groups living in close proximity, sharing only the attributes of their common status as redundant labor and common desperation for the sporadic, marginal employment available to them. Thus, the material basis for ethnically articulated conflict is established by the structural requirement of capitalism for a surplus population (see Smith 1981).

Peer groups which reinforce patterned delinquent behavior and those which reinforce conventional behavior are characterized by distinct structures of internal control that affect individual bonds and reinforce specific types of behavior. The internal structures of control within peer groups spring from control structures within the school and from differential opportunity structures in the social and economic environment of neighborhoods. Depending on the interaction between specific internal structures of control within delinquent groups and specific external structures of opportunity in the surrounding environment, the delinquent peer groups will produce either instrumental or violent patterned delinquent behavior.

Summary

Our theory focuses on the structures of control in several locations in the production and social reproduction processes. Our discussion of structures of control incorporates insights about micro-level processes from learning, strain, control, labeling, and integrated theories in criminology. We suggest that structures of control have parallel patterns associated with work, families, schools, and peer groups, and that those patterns form the mechanisms for the reproduction of the class structure. Expanding on insights about macro-level processes from conflict and radical criminology, we construct a structural-Marxist approach that views delinquency as a latent outcome of the social reproduction process in capitalism.

The structures of control tend to produce differential ideological bonds, depending on the individual's particular path through the control structures. We see ideological bonds as malleable, but they are usually sustained in the socialization process by somewhat similar experiences in control structures that individuals encounter along the path of socialization. The direction of socialization is initiated by the parents' location in workplace control structures, which are shaped by the historical interaction between competition among capitalists and the level of class struggle. These workplace control structures affect the structures of control within families. Children's initial bonds are shaped by family control relations and tend to set the child up for, or preclude placement in, specific control structures at school. School control structures create differential experiences of reward and punishment and reinforce or attenuate initial bonds. The juvenile is then open for recruitment to a variety of peer group experiences that are also shaped by structures of control among peers, which interact with differential opportunity structures in the surrounding community to produce specific patterns of peer group behavior. If patterned delinquent peer groups are available in the immediate social environment, a juvenile's structurally induced bond will open him up to, or insulate him from, entry into such peer relations. Entry into this type

of peer association continues the pattern of reinforcement toward more sustained delinquent behavior.

Figure 2 presents a general path model of delinquency production focusing on the individual's structurally conditioned experiences as he moves toward becoming a more serious, patterned delinquent. From our integrated structural-Marxist approach, we predict the following empirical relationships (numbered as in the figure).

Parents' *class position* (defined by ownership and control over the means and ends of production) is (*1*) negatively associated with parents' experience of coerciveness in *workplace control structures,* which (*2*) is positively associated with more alienated bonds in parents.

Alienated parental bonds (*3*) contribute to the development of more coercive *family control structures,* which (*4*) are positively related to more alienated initial bonds in juveniles.

Juveniles with alienated initial bonds (*5*) have a greater probability of being placed in more coercive *school control structures,* which (*6* and *7*) reinforce the juveniles' alienated bonds.

Juveniles' reinforced alienated bonds lead to (*8* and *9*) greater association with alienated peers, who form *peer group control structures* which, in interaction with (*10, 11,* and *12*) class-related, community, and neighborhood distributions of opportunities, create qualitatively different paths of delinquent development (*13* and *14*).

In the first path (leading from line *13*), the experience of coerciveness in peer group control relations mutually interacts (*15* and *16*) with juveniles' alienated bonds to (*17* and *18*) propel these juveniles into serious, patterned, *violent delinquent behavior.*

In the second path (leading from line *14*), the experience of remuneration from illegitimate sources (*19* and *20*) creates an alternative utilitarian control structure which mutually interacts (*21* and *22*) with newly formed calculative bonds to propel these juveniles (*23* and *24*) into serious, patterned, *instrumental delinquent behavior.*

CONCLUSION

Our structural-Marxist perspective precludes the type of social tinkering characteristic of the liberal-technocratic approach to problem solving. The apparent failure of liberal-technocratic approaches to have an impact on street crime and other social problems has left criminology (and the social sciences in general) in a theoretical vacuum. This vacuum could present an opportunity for creative reconceptualization and theory construction if assumptions concerning the sanctity of the basic social order were no longer taken for granted. However, in the absence of a rigorous critique of prevailing power relations, the theoretical vacuum can be a

543

Parents' Class Position

Experienced Coerciveness of Workplace Control Structure

Degree of Parents' Alienated Bonds

Coerciveness of Family Control Structures

Degree of Alienation of Juveniles' Initial Bonds

Experienced Coerciveness of School Control Structure

Juveniles' Reinforced Alienated Bonds

Degree of Association with Alienated Peers

Distribution of Legitimate Opportunities in the Community

Connection to Illegitimate Opportunities

Coercive Peer Control Structure

Reinforced Alienated Bonds

Patterned Violent Delinquency

Utilitarian Peer Control Structure

Reinforced Calculative Bonds

Patterned Instrumental Delinquency

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

Positive Association

Negative Association

FIG. 2.—General path model of patterned delinquency

precursor to a period of intellectual retrenchment and reaction, in which analytical thought is replaced by simplistic, yet ideologically compelling, appeals to our fears and ignorance about crime. The latter possibility is evident in the recent recommendations for repressive approaches to crime presented by President Reagan's Task Force on Violent Crime (Browning 1982). It is, therefore, imperative that social scientists counter this reactionary drift in social policies not through mere political rhetoric but through analytical and research skills.

Our perspective recognizes the structural limitations of changing one seemingly isolated aspect of society without a concurrent alteration of the core relationships associated with material production. The pivotal factor is the workplace struggle by subordinates for more equal and democratically controlled forms of the production relations that shape their lives and the lives of their families. The involvement in the struggle for changing the relations of production toward greater control by direct producers must be an essential aspect, along with community organizing and educational reform, of any program designed to counter crime and delinquency. Action, then, is not just technical but also political. Interestingly, this is precisely the implication of studies that report a significant decline in crime rates and gang violence concomitant with the political mobilization of the underclass (as suggested explicitly by Browning [1982] and implicitly by Erlanger [1982] and Short [1974]). Such conscious human intervention in the often destructive social reproduction process must take place at all points of that process; otherwise, our efforts are doomed to failure as the underlying structure transforms our intended efforts into outcomes that only reproduce the effects we had hoped to eliminate.

REFERENCES

Akers, Ronald L. 1977. *Deviant Behavior: A Social Learning Approach*. Belmont, Calif.: Wadsworth.

Althusser, Louis. 1971. *Lenin and Philosophy*. New York: Monthly Review Press.

———. 1977. *For Marx*. London: New Left Books.

Anderson, Charles H. 1974. *The Political Economy of Social Class*. Englewood Cliffs, N.J.: Prentice-Hall.

Anderson, Perry. 1980. *Arguments within English Marxism*. London: Verso.

Appelbaum, R. 1979. "Born-again Functionalism? A Reconsideration of Althusser's Structuralism." *Insurgent Sociologist* 9:18–33.

Baran, Paul, and Paul M. Sweezy. 1966. *Monopoly Capital*. New York: Monthly Review Press.

Barber, Theodore Xenophon, David S. Calverly, Albert Forgione, John D. McPeake, John F. Chaves, and Barbara Bowen. 1969. "Five Attempts to Replicate the Experimenter Bias Effect." *Journal of Consulting and Clinical Psychology* 33:1–6.

Bauer, D. 1981. "Why Big Business Is Firing the Boss." *New York Times Magazine* (March 8), pp. 22ff.

Becker, Howard S. 1963. *The Outsiders*. New York: Free Press.

Blackburn, Robin, ed. 1972. *Ideology in Social Science.* Glasgow: Fontana.

Blumer, Herbert. 1969. *Symbolic Interactionism: Perspective and Method.* Englewood Cliffs, N.J.: Prentice-Hall.

Bowles, Samuel, and Herbert Gintis. 1976. *Schooling in Capitalist America.* New York: Basic.

Braithwaite, John. 1981. "The Myth of Social Class and Criminality Reconsidered." *American Sociological Review* 46:36–57.

Braverman, Harry. 1974. *Labor and Monopoly Capital.* New York: Monthly Review Press.

Briar, Scott, and Irving Piliavan. 1970. "Delinquency, Situational Inducements, and Commitment to Conformity." Pp. 281–91 in *Society, Delinquency and Delinquent Behavior,* edited by Harwin L. Voss. Boston: Little, Brown.

Browning, Frank. 1982. "Nobody's Soft on Crime Anymore." *Mother Jones* 7:25ff.

Burgess, Robert L., and Ronald L. Akers. 1966. "A Differential Association-Reinforcement Theory of Criminal Behavior." *Social Problems* 14:128–47.

Burris, Val. 1979. "The Structuralist Influence in Marxist Theory and Research." *Insurgent Sociologist* 9:4–17.

Carnoy, Martin, and Henry M. Levin, eds. 1976. *The Limits of Educational Reform.* New York: McKay.

Chambliss, William J. 1964. "A Sociological Analysis of the Law of Vagrancy." *Social Problems* 11:67–77.

——. 1969. *Crime and the Legal Process.* New York: McGraw-Hill.

——, ed. 1973. *Sociological Readings in the Conflict Perspective.* Reading, Mass.: Addison-Wesley.

——. 1975. "Toward a Political Economy of Crime." *Theory and Society* 2:152–70.

——. 1979. "Contradictions and Conflicts in Law Creation." Pp. 3–27 in *Research in Law and Sociology.* Vol. 2. Edited by Steven Spitzer. Greenwich, Conn.: JAI.

Chiricos, Theodore G., and Gordon P. Waldo. 1975. "Socioeconomic Status and Criminal Sentencing: An Empirical Assessment of a Conflict Proposition." *American Sociological Review* 40:753–72.

Cloward, Richard A., and Lloyd E. Ohlin. 1960. *Delinquency and Opportunity.* New York: Free Press.

Coburn, David, and Virginia L. Edwards. 1976. "Job Control and Child-rearing Values." *Canadian Review of Sociology and Anthropology* 13:337–44.

Cohen, Albert K. 1955. *Delinquent Boys.* Glencoe, Ill.: Free Press.

Colvin, Mark. 1981. "The Contradictions of Control: Prisons in Class Society." *Insurgent Sociologist* 10/11:33–46.

Conger, Rand D. 1976. "Social Control and Social Learning Models of Delinquent Behavior: A Synthesis." *Criminology* 14:17–40.

Crime and Social Justice Collective. 1976. "The Politics of Street Crime." *Crime and Social Justice* 5:1–4.

Dannefer, Dale, and Russell K. Schutt. 1982. "Race and Juvenile Justice Processing in Court and Police Agencies." *American Journal of Sociology* 87:1113–32.

Davis, Mike. 1981. "The New Right's Road to Power." *New Left Review* 128:28–49.

Davis, Nanette J. 1972. "Labeling Theory in Deviance Research: A Critique and Reconsideration." *Sociological Quarterly* 13:447–74.

——. 1975. *Sociological Constructions of Deviance.* Dubuque, Iowa: Brown.

Douglas, Jack D. 1970. *American Social Order.* New York: Free Press.

Edwards, Richard. 1979. *Contested Terrain: The Transformation of the Workplace in the Twentieth Century.* New York: Basic.

Einstein, Albert, and Leopold Infeld. 1938. *The Evolution of Physics.* New York: Simon & Schuster.

Elliot, Delbert S., and Suzanne S. Ageton. 1980. "Reconciling Race and Class Differences in Self-reported and Official Estimates of Delinquency." *American Sociological Review* 45:95–110.

Elliot, Delbert S., Suzanne S. Ageton, and Rachelle J. Canter. 1979. "An Integrated Theoretical Perspective on Delinquent Behavior." *Journal of Research in Crime and Delinquency* 16:3–27.

Empey, LaMar T., and Steven G. Lubeck. 1971. *Explaining Delinquency.* Lexington, Mass.: Heath.

Erickson, Maynard L., and LaMar T. Empey. 1965. "Class Position, Peers and Delinquency." *Sociology and Social Research* 49:268–82.

Erickson, Maynard L., and Gary F. Jensen. 1977. "Delinquency Is Still Group Behavior! Toward Revitalizing the Group Premise in the Sociology of Deviance." *Journal of Criminal Law and Criminology* 68:262–73.

Erlanger, Howard S. 1974. "Social Class and Corporal Punishment in Child-rearing: A Reassessment." *American Sociological Review* 39:92–107.

———. 1980. "The Empirical Status of the Subculture of Violence Thesis." Pp. 192–204 in *Criminal Behavior,* edited by Delos H. Kelly. New York: St. Martin's.

———. 1982. "Machismo and Gang Violence." Pp. 185–94 in *Contemporary Criminology,* edited by Leonard D. Savitz and Norman Johnston. New York: Wiley.

Etzioni, Amitai. 1970. "Compliance Theory." Pp. 103–26 in *The Sociology of Organizations,* edited by Oscar Grusky and George A. Miller. New York: Free Press.

Foster, Jack D., Simon Dinitz, and Walter C. Reckless. 1972. "Perceptions of Stigma Following Public Intervention for Delinquent Behavior." *Social Problems* 20:202–9.

Franklin, Jack I., and Joseph E. Scott. 1970. "Parental Values: An Inquiry into Occupational Setting." *Journal of Marriage and the Family* 32:406–9.

Garfinkel, Harold, ed. 1967. *Studies in Ethnomethodology.* Englewood Cliffs, N.J.: Prentice-Hall.

Gecas, Viktor. 1979. "The Influence of Social Class on Socialization." Pp. 365–404 in *Contemporary Theories about the Family,* edited by Wesley R. Burr, Ruben Hill, F. Ivan Nye, and Ira L. Reiss. Vol. 1. New York: Free Press.

Gecas, Victor, and F. Ivan Nye. 1974. "Sex and Class Differences in Parent-Child Interaction: A Test of Kohn's Hypothesis." *Journal of Marriage and the Family* 36:742–49.

Gephart, William J. 1970. "Will the Real Pygmalion Please Stand Up?" *American Educational Research Journal* 7:473–74.

Gibbons, Don C. 1976. *Delinquent Behavior.* 2d ed. Englewood Cliffs, N.J.: Prentice-Hall.

———. 1979. *The Criminological Enterprise: Theories and Perspectives.* Englewood Cliffs, N.J.: Prentice-Hall.

Glimenez, Martha E. 1982. "The Oppression of Women: A Structuralist Marxist View." Pp. 292–324 in *Structural Sociology,* edited by Ino Rossi. New York: Columbia University Press.

Glueck, Sheldon, and Eleanor Glueck. 1950. *Unraveling Juvenile Delinquency.* New York: Commonwealth Fund.

Godelier, Maurice. 1972. *Rationality and Irrationality in Economics.* New York: Monthly Review Press.

Goffman, Erving. 1963. *Stigma.* Englewood Cliffs, N.J.: Prentice-Hall.

Gouldner, Alvin W. 1968. "The Sociologist as Partisan: Sociology and the Welfare State." *American Sociologist* 3:103–16.

Greenberg, David F. 1976. "On One-Dimensional Marxist Criminology." *Theory and Practice* 3:611–21.

———. 1977. "Delinquency and the Age Structure of Society." *Contemporary Crises* 1:189–223.

———, ed. 1981. *Crime and Capitalism.* Palo Alto, Calif.: Mayfield.

Hepburn, John R. 1976. "Testing Alternative Models of Delinquency Causation." *Journal of Criminal Law and Criminology* 67:450–60.

Hewitt, John D. 1977. "The Effects of Individual Resources in Judicial Sentencing." *Public Data Use* 5:28–48.

Hills, Stuart L. 1971. *Crime, Power and Morality.* Scranton, Pa.: Chandler.

Hindelang, Michael J. 1973. "Causes of Delinquency: A Partial Replication and Extension." *Social Problems* 20:471–87.

Hirschi, Travis. 1969. *Causes of Delinquency.* Berkeley and Los Angeles: University of California Press.

Hirschi, Travis, and Michael J. Hindelang. 1977. "Intelligence and Delinquency: A Revisionist Review." *American Sociological Review* 42:571–87.

Humphries, Drew, and David F. Greenberg. 1981. "The Dialectics of Crime Control." Pp. 209–54 in Greenberg., ed., 1981.

Jencks, Christopher, Marshall Smith, Henry Acland, Mary Jo Bane, David Cohen, Herbert Gintis, Barbara Heyns, and Stephan Michelson. 1972. *Inequality: A Reassessment of the Effect of Family and Schooling in America*. New York: Basic.

Jensen, Gary F., and Dean G. Rojek. 1980. *Delinquency: A Sociological View*. Lexington, Mass.: Heath.

Jessor, Richard, Theodore D. Graves, Robert C. Hanson, and Shirley L. Jessor. 1968. *Society, Personality and Deviant Behavior*. New York: Holt, Rinehart & Winston.

Johnson, Richard E. 1979. *Juvenile Delinquency and Its Origins: An Integrated Theoretical Approach*. Cambridge: Cambridge University Press.

Kelly, Delos H. 1974. "Track Position and Delinquent Involvement: A Preliminary Analysis." *Sociology and Social Research* 58:380–86.

Kitsuse, John I., and David C. Dietrick. 1959. "Delinquent Boys: A Critique." *American Sociological Review* 24:208–15.

Kohn, Melvin L. 1976. "Social Class and Parental Values: Another Confirmation of the Relationship." *American Sociological Review* 41:538–45.

———. 1977. *Class and Conformity*. Chicago: University of Chicago Press.

Kohn, Melvin L., and Carmi Schooler. 1973. "Occupational Experience and Psychological Functioning: An Assessment of Reciprocal Effects." *American Sociological Review* 38:97–118.

Komarovsky, Mirra. 1962. *Blue Collar Marriage*. New York: Random House.

Laclau, Ernesto. 1979. *Politics and Ideology in Marxist Theory*. London: Verso.

Leacock, Eleanor. 1969. *Teaching and Learning in City Schools*. New York: Basic.

Lemert, Edwin M. 1972. *Human Deviance, Social Problems and Social Control*. 2d ed. Englewood Cliffs, N.J.: Prentice-Hall.

Levison, Andrew. 1974. *The Working Class Majority*. New York: Penguin.

Liazos, A. 1972. "The Poverty of the Sociology of Deviance: Nuts, Sluts and Perverts." *Social Problems* 20:103–20.

Lofland, John. 1969. *Deviance and Identity*. Englewood Cliffs, N.J.: Prentice-Hall.

Maccoby, Eleanor E., Joseph P. Johnson, and Russell M. Church. 1958. "Community Integration and the Social Control of Juvenile Delinquency." *Journal of Social Issues* 14:38–51.

McCord, William, Joan McCord, and Irving Zola. 1959. *Origins of Crime*. Montclair, N.J.: Patterson-Smith.

Mahoney, Ann Rankin. 1974. "The Effect of Labeling upon Youths in the Juvenile Justice System: A Review of the Evidence." *Law and Society Review* 8:583–614.

Manning, Peter. 1975. "Deviance and Dogma." *British Journal of Criminology* 15:1–20.

Marx, Karl. 1963. *The Eighteenth Brumaire of Louis Bonaparte*. New York: International Publishers.

———. 1977. *Capital*, vol. 1. New York: Vintage.

———. 1981. *Capital*, vol. 3. New York: Vintage.

Marx, Karl, and Frederick Engels. 1967. *The German Ideology*. New York: International Publishers.

Matza, David. 1964. *Delinquency and Drift*. New York: Wiley.

———. 1969. *Becoming Deviant*. Englewood Cliffs, N.J.: Prentice-Hall.

Mead, George H. 1934. *Mind, Self and Society*. Chicago: University of Chicago Press.

Merton, Robert K. 1938. "Social Structure and Anomie." *American Sociological Review* 3:672–82.

Miller, Walter B. 1958. "Lower Class Culture as a Generating Milieu of Gang Delinquency." *Journal of Social Issues* 14:5–19.

Mills, C. Wright. 1943. "The Professional Ideology of Social Pathologists." *American Journal of Sociology* 49:165–80.

Montgomery, David. 1979. *Workers' Control in America*. Cambridge: Cambridge University Press.

Noble, David. 1977. *America by Design*. New York: Oxford University Press.

Nye, F. Ivan. 1958. *Family Relationships and Delinquent Behavior*. Westport, Conn.: Greenwood.

O'Connor, James. 1973. *The Fiscal Crisis of the State.* New York: St. Martin's.

Oster, Gerry. 1979. "A Factor Analytic Test of the Theory of the Dual Economy." *Review of Economics and Statistics* 61:33–39.

Parker, Richard. 1972. *The Myth of the Middle Class.* New York: Harper Colophon.

Pearlin, Leonard I. 1970. *Class Context and Family Relations: A Cross-national Study.* Boston: Little, Brown.

Piliavan, Irving, and Scott Briar. 1964. "Police Encounters with Juveniles." *American Journal of Sociology* 70:206–14.

Piven, Frances Fox, and Richard A. Cloward. 1971. *Regulating the Poor: The Functions of Public Welfare.* New York: Vintage.

Platt, Tony. 1974. "Prospects for a Radical Criminology in the United States." *Crime and Social Justice* 1:2–10.

————. 1978. "Street Crime: A View from the Left." *Crime and Social Justice* 9:26–34.

Post, Charles. 1982. "The American Road to Capitalism." *New Left Review* 133:30–51.

Poulantzas, Nicos. 1978a. *Classes in Contemporary Capitalism.* London: Verso.

————. 1978b. *Political Power and Social Classes.* London: Verso.

————. 1980. *State, Power, Socialism.* London: Verso.

Quinney, Richard. 1970. *The Social Reality of Crime.* Boston: Little, Brown.

————. 1974a. *Criminal Justice in America.* Boston: Little, Brown.

————. 1974b. *Critique of Legal Order: Crime Control in Capitalist Society.* Boston: Little, Brown.

————. 1975. *Criminology.* Boston: Little, Brown.

————. 1977. *Class, State and Crime.* New York: Longman.

Reckless, Walter C., and Simon Dinitz. 1970. "Pioneering with Self-Concept as a Vulnerability Factor in Delinquency." Pp. 175–84 in *Society, Delinquency and Delinquent Behavior,* edited by Harwin L. Voss. Boston: Little, Brown.

Reiss, Albert J. 1951. "Delinquency as the Failure of Personal Social Controls." *American Sociological Review* 16:196–207.

Rist, Ray C. 1970. "Student Social Class and Teacher Expectations: The Self-fulfilling Prophecy in Ghetto Education." *Harvard Educational Review* 40:411–51.

Rock, Paul. 1974. "The Sociology of Deviance and Conceptions of Moral Order." *British Journal of Sociology* 14:139–49.

Rodman, Hyman. 1968. "Controversies about Lower Class Culture: Delinquency and Illegitimacy." *Canadian Review of Sociology and Anthropology* 5:254–62.

Rosenthal, Robert. 1973. "Pygmalion Effect Lives." *Psychology Today* 7:56–63.

Rosenthal, Robert, and Lenore Jacobson. 1968. *Pygmalion in the Classroom.* New York: Holt, Rinehart & Winston.

Rubin, Lillian B. 1976. *Worlds of Pain.* New York: Basic.

Rubovits, P. C., and A. L. Maehr. 1971. "Pygmalion Analyzed: Toward an Explanation of the Rosenthal-Jacobson Findings." *Journal of Personality and Social Psychology* 19:197–203.

Schaefer, Walter E., Carol Olexa, and Kenneth Polk. 1972. "Programmed for Social Class: Tracking in High School." Pp. 33–54 in *Schools and Delinquency,* edited by Kenneth Polk and Walter E. Schaefer. Englewood Cliffs, N.J.: Prentice-Hall.

Schur, Edwin M. 1971. *Labeling Deviant Behavior.* New York: Harper & Row.

Schwendinger, Herman, and Julia Schwendinger. 1970. "Defenders of Order or Guardians of Human Rights?" *Issues in Criminology* 7:72–81.

————. 1976a. "Delinquency and the Collective Varieties of Youth." *Crime and Social Justice* 5:7–25.

————. 1976b. "Marginal Youth and Social Policy." *Social Problems* 24:84–91.

Sellin, Thorsten. 1938. *Culture Conflict and Crime.* New York: Social Science Research Council.

Short, James F. 1958. "Differential Association with Delinquent Friends and Delinquent Behavior." *Pacific Sociological Review* 1:20–25.

————. 1974. "Youth, Gangs and Society: Micro- and Macrosociological Processes." *Sociological Quarterly* 15:3–19.

Short, James F., Ramon Rivera, and Ray A. Tennyson. 1965. "Perceived Opportunities, Gang Membership and Delinquency." *American Sociological Review* 30:56–57.

Simpson, George E., and J. Milton Yinger. 1965. *Racial and Cultural Minorities*. New York: Harper & Row.

Skinner, B. F. 1953. *Science and Human Behavior*. New York: Macmillan.

Smith, Joan. 1981. *Social Issues and the Social Order: The Contradictions of Capitalism*. Cambridge, Mass.: Winthrop.

Sparks, Richard. 1980. "A Critique of Marxist Criminology." Pp. 159–210 in *Crime and Justice*, edited by Norval Morris and Michael Tonry. Vol. 2. Chicago: University of Chicago Press.

Spitzer, Steven. 1975. "Toward a Marxian Theory of Deviance." *Social Problems* 22:638–51.

———, ed. 1979. *Research in Law and Sociology*, vol. 2. Greenwich, Conn.: JAI.

———. 1980. "'Left-wing' Criminology—an Infantile Disorder?" Pp. 169–90 in *Radical Criminology: The Coming Crises*, edited by James Inciardi. Beverly Hills, Calif.: Sage.

Spring, Joel H. 1972. *Education and the Rise of the Corporate State*. Boston: Beacon.

Steinert, Heinz. 1978. "Can Socialism Be Advanced by Radical Rhetoric and Sloppy Data? Some Remarks on Richard Quinney's Latest Output." *Contemporary Crises* 2:303–13.

Strauss, Murray A., Richard J. Gelles, and Suzanne K. Steinmetz. 1980. *Behind Closed Doors: Violence in the American Family*. New York: Doubleday.

Sutherland, Edwin H. 1947. *Principles of Criminology*, 4th ed. Philadelphia: Lippincott.

———. 1949. *White Collar Crime*. New York: Holt, Rinehart & Winston.

Sutherland, Edwin H., and Donald R. Cressey. 1960. *Principles of Criminology*, 6th ed. Philadelphia: Lippincott.

———. 1978. *Criminology*, 10th ed. Philadelphia: Lippincott.

Sykes, Gresham M. 1974. "The Rise of Critical Criminology." *Journal of Criminal Law and Criminology* 65:206–13.

Taylor, Ian, Paul Walton, and Jock Young. 1973. *The New Criminology*. New York: Harper Torchbooks.

Therborn, Goran. 1980. *The Ideology of Power and the Power of Ideology*. London: Verso.

Thomas, C. W., and J. M. Hyman. 1978. "Compliance Theory, Control Theory, and Juvenile Delinquency." Pp. 73–90 in *Crime, Law and Sanctions*, edited by Marvin D. Krohn and Ronald L. Akers. Beverly Hills, Calif.: Sage.

Thompson, E. P. 1978. *The Poverty of Theory and Other Essays*. New York: Monthly Review Press.

Thompson, Randall J., and Matthew T. Zingraff. 1981. "Detecting Sentencing Disparity: Some Problems and Evidence." *American Journal of Sociology* 86:869–80.

Tittle, Charles R., Wayne J. Villemez, and Douglas A. Smith. 1978. "The Myth of Social Class and Criminality: An Empirical Assessment of the Empirical Evidence." *American Sociological Review* 43:643–56.

Toby, Jackson. 1957. "Social Disorganization and Stake in Conformity: Complementary Factors in the Predatory Behavior of Hoodlums." *Journal of Criminal Law, Criminology and Police Science* 48:12–17.

Tolbert, Charles, Patrick M. Horan, and E. M. Beck. 1980. "The Structure of Economic Segmentation: A Dual Economy Approach." *American Journal of Sociology* 85:1095–1116.

Turk, Austin T. 1969. *Criminality and Legal Order*. Chicago: Rand McNally.

———. 1975. "Prospects and Pitfalls for Radical Criminology: A Critical Response to Platt." *Crime and Social Justice* 4:41–42.

Vold, George. 1958. *Theoretical Criminology*. New York: Oxford University Press.

Voss, Harwin L. 1964. "Differential Association and Reported Delinquent Behavior: A Replication." *Social Problems* 12:78–85.

West, Donald J., and David P. Farrington. 1973. *Who Becomes Delinquent?* London: Heinemann.

Whaley, Donald L., and Richard W. Malott. 1969. *Elementary Principles of Behavior*. Kalamazoo, Mich.: Behaviordelia.

Wiatrowski, Michael D., Stephen Hansell, Charles R. Massey, and David L. Wilson. 1982. "Curriculum Tracking and Delinquency." *American Sociological Review* 47:151–60.

Wolfgang, Marvin E., ed. 1968. *Crime and Culture: Essays in Honor of Thorsten Sellin*. New York: Wiley.

Wright, Eric Olin. 1978. *Class, Crisis and State*. London: New Left Books.

———. 1979. *Class Structure and Income Determination*. New York: Academic Press.

Wright, James D., and Sonia R. Wright. 1976. "Social Class and Parental Values for Children: A Partial Replication and Extension of the Kohn Thesis." *American Sociological Review* 41:527–37.

Yablonsky, Lewis. 1962. "The Delinquent Gang as a Near Group." *Social Problems* 7:108–17.

———. 1966. *The Violent Gang*. Baltimore: Penguin.

Age and the Explanation of Crime[1]

Travis Hirschi
University of Arizona

Michael Gottfredson
University of Illinois at Urbana-Champaign

One of the few facts agreed on in criminology is the age distribution of crime. This fact has been used to criticize social theories of crime causation, to provide the foundation for other theories, to justify recent emphases on career criminals, and to support claims of superiority for longitudinal designs in criminological research. In the present paper, we argue that the age distribution of crime is sufficiently invariant over a broad range of social conditions that these uses of the age distribution are not justified by available evidence.

According to a recent criminology textbook (Allen et al. 1981, p. 235), age is the easiest fact about crime to study. In one sense, the statement is true: the age of the offender is routinely recorded, and age distributions of crime covering a variety of contexts over a long period are not hard to find.[2] As a result, no fact about crime is more widely accepted by criminologists. Virtually all of them, of whatever theoretical persuasion, appear to operate with a common image of the age distribution. This distribution thus represents one of the brute facts of criminology. Still, the statement that age is an easy fact to study is decidedly misleading. When attention shifts to the meaning or implications of the relation between age and crime, that relation easily qualifies as the most difficult fact in the field. Efforts to discern the meaning of the large amount of research on the topic in terms supplied by those doing the research have

[2] Most of the available data on age and crime are "official" data—i.e., data on arrestees or prisoners. They are, therefore, subject to traditional criticisms of official data: they may reflect biased enforcement rather than the behavior of offenders. The logical forms such bias could take are virtually endless. For example, juveniles may be more likely than adults to be arrested either because they commit a larger portion of their offenses in groups and groups are more easily apprehended, or because they are less skilled in evasion. Or juveniles may be less vulnerable to arrest than adults because of leniency. Empirical examination of such biases, with data from self-report and victimization surveys, suggests that the biases do not account for the relation between age and crime (see, e.g., Empey 1978).

turned out to be futile (e.g., Wootton 1959, chap. 5; see also Wolfgang, Figlio, and Sellin 1972, p. 105), as have efforts to explain the relation in statistical terms (Rowe and Tittle 1977).[1]

Faced with this intransigent fact, the response in criminology has been generally scientific and logical. Theorists are frequently reminded that their explanations of crime must square with the age distribution, and theories are often judged by their ability to deal with "maturational reform," "spontaneous remission," or the "aging-out" effect (Matza 1964; Hirschi 1969; Empey 1978; Siegal and Senna 1981). Although some theories fare better than others when the age criterion is invoked, no theory that focuses on differences between offenders and nonoffenders avoids altogether the complaint that it provides an inadequate explanation of the age distribution. Given the persuasiveness of the age criticism of traditional theories, it is not surprising to find recent explanations of crime explicitly tailored to fit the accepted variability in crime by age (Matza 1964; Greenberg 1979; Trasler 1980). In fact, there is reason to believe that age could replace social class as the master variable of sociological theories of crime (see Empey 1978; Glaser 1978; Greenberg 1979).

On the research side, the age effect has been instrumental in the rise of the longitudinal study to its current status as the preferred method of criminological research. The major studies of the past decade, including several still under way, have used this design (Wolfgang et al. 1972; West and Farrington 1973; Elliott, Ageton, and Huizanga 1978; McCord 1979; Wadsworth 1979). This research emphasis gains much of its attractiveness from the association between age and such concepts as "career criminal," "recidivism," and "desistance," all of which are thought to be of considerable theoretical and practical import and all of which are thought to require, by definition, longitudinal designs for their study.

Given the increasing role of age in criminological theory and research (e.g., Elliott, Ageton, and Canter 1979; Greenberg 1979; Farrington 1979; Matza 1964; Petersilia 1980; Zimring 1981) and the widely accepted critique of sociological theories on the basis of the age effect (e.g., Matza 1964; Hirschi 1969; Empey 1978; Siegel and Senna 1981), it seems to us that those in the field should consider the possibility that current conceptions of the age effect and its implications for research and theory are misguided. To that end, in this paper we advance and attempt to defend

[1] The Rowe and Tittle article is a thorough attempt to account for the age effect in a large sample of respondents 15 and older whose criminality was measured by self-reports. Although the authors report isolating a segment of their sample (of unknown size) in which there is no relation between age and crime, we believe this group is composed largely of respondents whose self-reports are manifestly unreliable. It is consistent with our hypothesis that the relation between age and crime is stronger in the "reliable" portion of the sample than in any of the many other subgroups Rowe and Tittle isolate.

the following theses: (1) the age distribution of crime is invariant across social and cultural conditions; (2) theories of crime that do not explicitly attend to age have no logical or empirical obligation to do so and should not be judged by their apparent ability or inability to account for the age effect; (3) the age distribution of crime cannot be accounted for by any variable or combination of variables currently available to criminology; (4) explanations focusing explicitly on the age effect must be compatible with an apparently direct effect of age on crime; (5) the conceptual apparatus that has grown up around the age effect is largely redundant or misleading; (6) identification of the causes of crime at any age may suffice to identify them at other ages as well; if so, cohort or other longitudinal designs are not necessary for the proper study of crime causation.

We recognize the difficulty in establishing some of our theses, especially those that deny either the significance of variability not yet investigated or the power of explanations not yet advanced. Nevertheless, we find nothing in the available research literature inconsistent with our position, and we find a good deal to support it. Furthermore, we consider our specific theses to be logically connected to such an extent that we are obligated to explore each and all of them; were we to discard the difficult or inconvenient issues, we would in effect grant the untenability of our entire position, something we are not now prepared to do. If currently popular approaches to the age distribution are inadequate, there seems little reason to pursue them along traditional lines without considering an alternative position.[4]

1. THE AGE EFFECT IS INVARIANT

Theoretical and textbook discussions of the age effect often presuppose or flatly assert (the former is more common) variations in this effect over time, place, demographic subgroups, or type of crime (Empey 1978, p. 391; Jensen and Rojek 1980, pp. 70–71; Reckless 1973, p. 81; Glaser 1978; Allen et al. 1981, pp. 234–35). Typically, the current age distribution of crime in the United States as revealed by the Uniform Crime Reports (UCR) is shown and the reader is left with the impression that this distribution is only one of many such distributions revealed by research.

---

[4] Collectively, reviewers have recommended that we seriously qualify or abandon the majority of our theses, including the argument that the age effect is invariant, which is the key to the remainder. The general complaint is that our assertions are "too strong," "sweeping," or "unsubstantiated." We believe that our assertions are derived from the evidence, although we grant that in many cases they go beyond it. If we are wrong, it should not be too hard to show that we are wrong. The risk to the field of considering our hypotheses would therefore appear to be minimal.

Time and Place

Figures 1, 2, and 3 show three age distributions of criminality:[5] one from England and Wales in 1842–44 as reported by Neison in 1857, another from England in 1908 as reported by Goring in 1913, and another from the most recently available UCR for the United States (U.S. Department of Justice 1979). Looking at one of these distributions, Goring (1913) concluded that the age distribution of crime conformed to a "law of nature."[6] The similarity between the three distributions is sufficient to suggest that little or nothing has happened to Goring's law of nature since he first discovered it. The shape or form of the distribution has remained virtually unchanged for about 150 years.[7] Recent data, the basis for many assertions of variability in the age distribution, force the same conclusion: ". . . while population arrest rates have changed in absolute magnitude over time (almost doubling between 1965 and 1976), the same pattern has persisted for the relative magnitudes of the different age groups, with fifteen to seventeen year-olds having the highest arrest rates per population of any age group" (Blumstein and Cohen 1979, p. 562).

We do not know how England and Wales in the 1840s differed from the United States in the 1980s. Presumably the differences are large across a variety of relevant dimensions. We do know, however, that in the 1960s, the age distribution of delinquency in Argentina (DeFleur 1970, p. 131) was indistinguishable from the age distribution in the United States, which was in turn indistinguishable from the age distribution of delinquency in England and Wales at the same time (McClintock and Avison 1968). If the form of the age distribution differs from time to time and from place to place, we have been unable to find evidence of this fact.

[5] The figures presented in this paper are only illustrative of many similar figures that could be constructed from the literature. The following sources contain age-crime distributions for Belgium, Sweden, West Germany, France, Italy, Russia, and Japan, respectively, at various times in this century, largely since World War II: Swedish National Central Bureau of Statistics 1980; Jacquart, n.d.; Mannheim 1965; Rangol 1962; Besson 1961; Ponti 1962; Callcott 1935.

[6] Goring contrasted the observed distribution of age at first conviction among habitual criminals with two theoretical probability distributions (Pearson's Type I and III) and found a poor fit. However, he concluded that exclusion of the youngest age groups (10–20-year-olds) produced a close fit between observed and expected Type I distributions. Arguing that interference by artificial agencies produces the initial incompatibility between the two distributions ("the efforts of the law to postpone its penalties in the case of juveniles"), Goring concluded that "the age-frequencies at first conviction of habitual criminals . . . obey natural laws of frequency . . ." (Goring 1913, p. 211).

[7] This conclusion is supported by data from Quetelet (1969), U.S. prison statistics (U.S. Department of Commerce, Bureau of the Census 1893; Best 1930), and by the UCR over the period in which age statistics have been published (1930s to 1980). Although data from the Soviet Union are sparse, available sources indicate that they are probably comparable: "Soviet writers generally assert that delinquency is concentrated in the 'troublesome' years between 15 and 17" (Conner 1970, p. 286). See also the studies cited in n. 5 above.

Demographic Groups

Most discussions of the age distribution in a theoretical context assume important differences for demographic subgroups. Textbooks often compare rates of increase in crime for boys and girls for particular offenses, thus suggesting considerable flexibility in the age distribution by sex. "Age-of-onset" studies easily suggest that, say, black offenders "start earlier" than white offenders; such a suggestion gives the impression that the age distribution of crime varies across ethnic or racial groups (see, e.g., Wolfgang et al. 1972, p. 131). Figures 4 (sex) and 5 (race) show that such suggestions tend to obscure a basic and persistent fact: available data suggest that the age-crime relation is invariant across sex and race.[8]
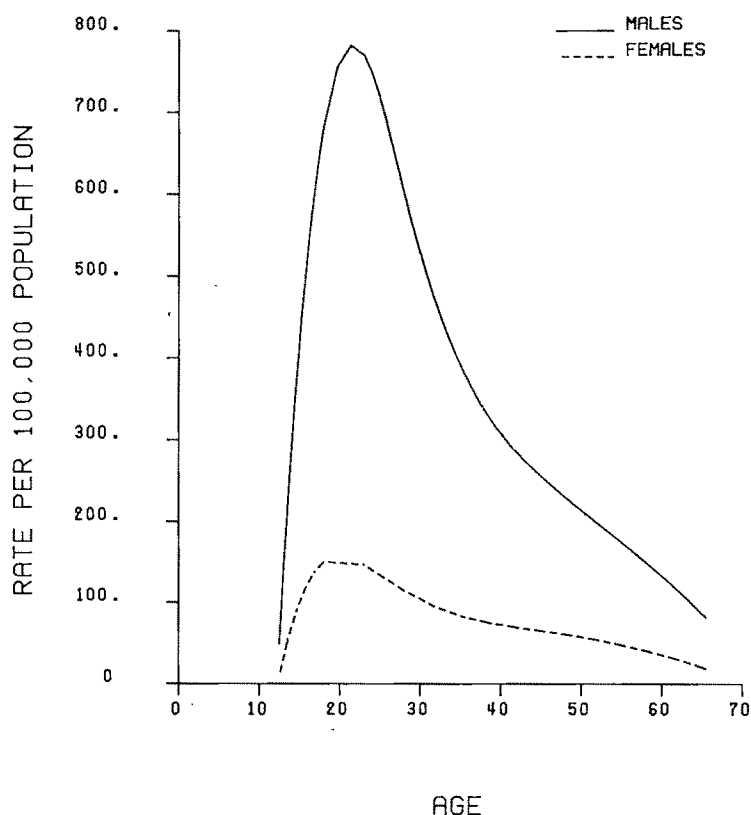
FIG. 1.—Criminal offenders in England and Wales, 1842, 1843, 1844, by age and sex. Rates per 100,000 population. (Source of data: Neison [1857, pp. 303–4].)

[8] See also fig. 1, where the Neison (1857) data for males and females are drawn to the same scale. At first glance, the figure suggests that the age distribution of crime for females is "flatter" than that for males, an impression sometimes reported in the literature. The fact is that it is not flatter. On the contrary, except for differences in level, the curves are virtually the same.

556

Type of Crime

A consistent difference in the age distribution of person and property offenses appears to be well-established, at least for official data. In such data, person crimes peak later than property crimes, and the rate declines more slowly with age. The significance of this fact for theories of criminality is, however, problematic. For one thing, self-report data do not support the distinction between person and property offenses; they show instead that both types of offense peak at the same time (see, e.g., Elliott et al. 1978) and decline at the same rate with age (Tittle 1980). The peak years for person and property offenses in self-report data are the mid-teens, which are also the peak years for property offenses in official data. In contrast, person offenses in official data peak in the late teens or early twenties.

AGE
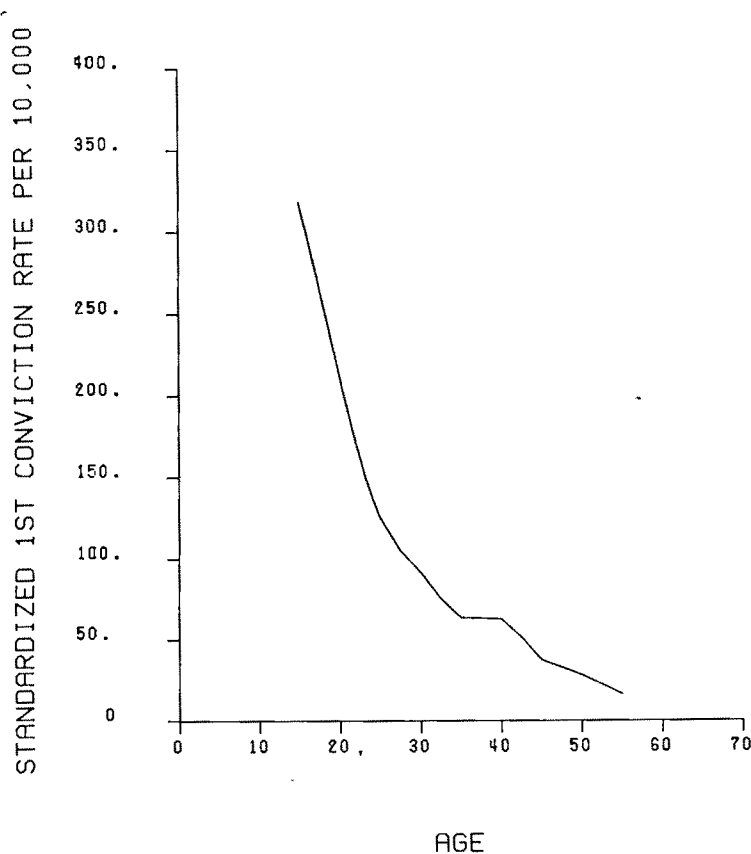
FIG. 2.—Frequencies of age of criminals at first conviction as a percentage of the frequencies of age in the general population. Males over 15 years, England, 1908. (Source of data: Goring [1913, pp. 201–2].)

If the self-report results are taken as indicative of the level of criminality, the difference in the peak years for person and property offenses in official data may be accounted for by age-related differences in the *consequences* of person and property crimes. One of these differences lies in the seriousness of offenses. Wolfgang and his colleagues report that "injury seriousness scores advance dramatically at each offense rank number," while the increase in seriousness for theft offenses is negligible (1972, p. 171). Offense rank is correlated with age (as a group, second offenders are older than first offenders). It should follow that age is positively correlated with the seriousness of injury offenses but not with the seriousness of theft offenses. By extension (and this is consistent with everyday observation), "injury" offenses by the very young are unlikely to be sufficiently serious to attract the attention of officials. Indeed, as long ago as 1835, Quetelet (1969) presented data on the correlation between physical strength and
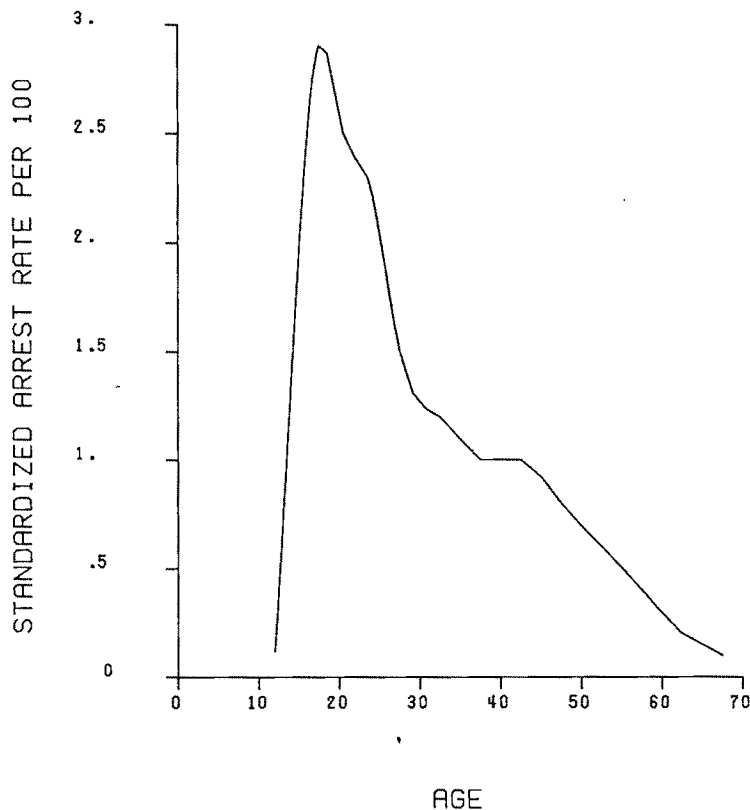
FIG. 3.—Age distribution of persons arrested in the United States for all offenses, standardized for their representation in general population, 1977. (Source of data: U.S. Department of Justice [1979, p. 171]. N.B.: Data are approximate.)

age alongside data on the age distribution of crime, the idea being that some crimes appear only when the strength necessary to inflict injury or coerce others has been attained. Apparently, the tendency to commit criminal acts, as reflected in theft offenses, however measured, and in violent offenses, as measured by self-reports, peaks before the physical ability necessary for serious violent offenses. The peak age for person offenses is thus a consequence of the confluence of the "tendency" and "ability" curves. Since strength continues to increase after the peak age of criminality has been reached, the person-crime curve declines from a later point. For a brief period, increases in the dangerousness of offenders more than offset their declining tendency to commit offenses.

The slower decline of person offenses in official data may reflect the fact that a greater proportion of such offenses involve primary group
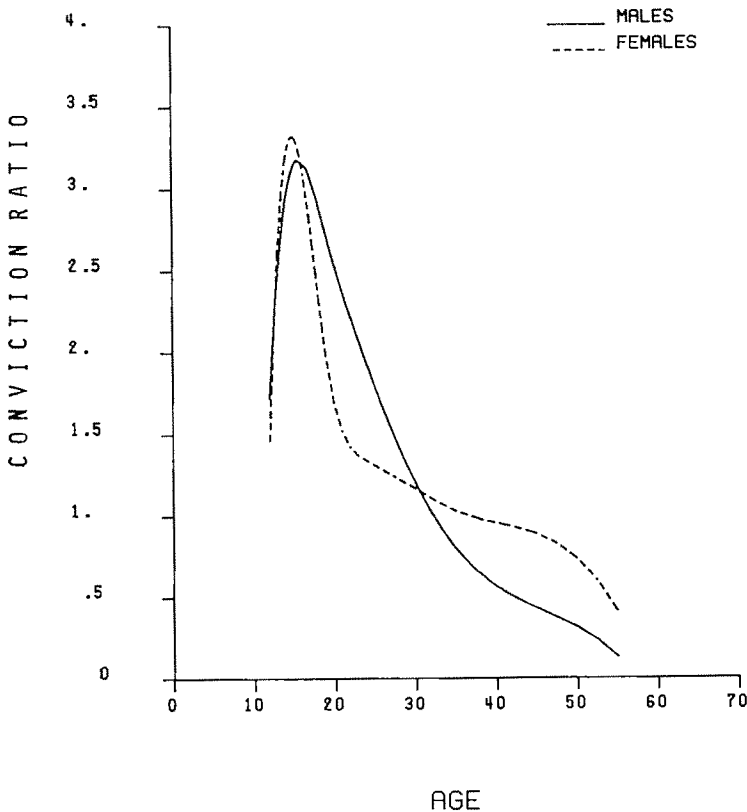
FIG. 4.—Age distribution of persons found guilty of indictable offenses, standardized by sex, England and Wales, 1965. Conviction ratio: percentage of offenders in age group divided by percentage of population in age group. (Source of data: McClintock and Avison [1968, p. 170].)

conflicts. Primary group conflicts may be assumed to be relatively constant over the age span and to produce a relatively stable number of assaultive offenses during the period of capability (i.e., among those neither very young nor very old). If these offenses were subtracted from the total number of person offenses, the form of the age curve for person offenses would approximate more closely that for property offenses. These speculations are consistent with the self-report finding of no difference between person and property crimes with respect to the long-term effects of age (Tittle 1980, p. 92).

Since our thesis is that the age effect is invariant across social and cultural conditions, it may appear that our explanation of the apparent difference between person and property crimes requires modification of our thesis. Actually, in some social conditions, the effects of age may be muted. As people retreat into the primary group context with increasing
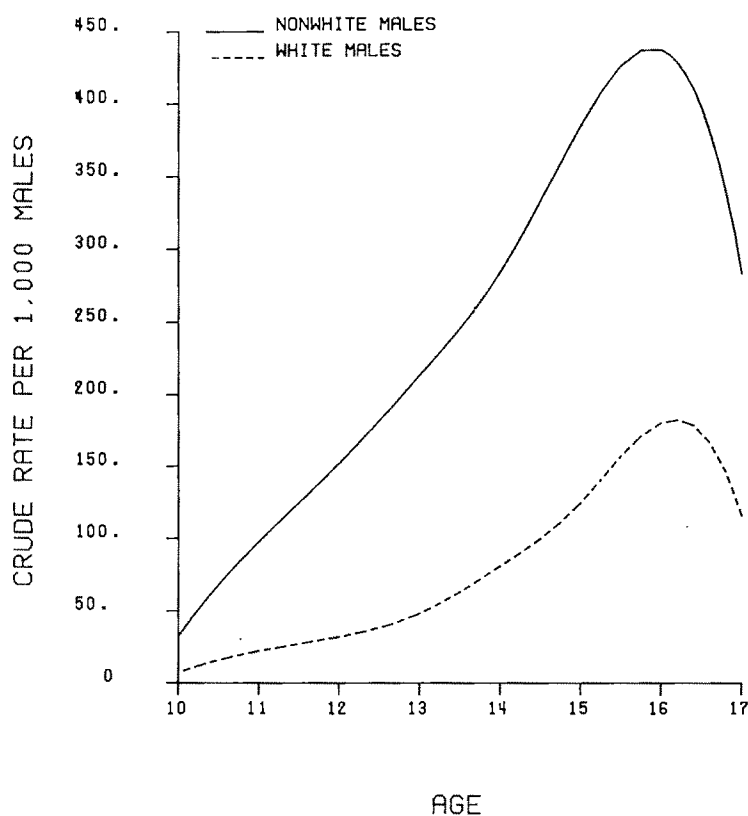
FIG. 5.—Delinquency rates by race and age. (Source of data: Wolfgang et al. [1972, p. 109].)

age, the relatively rare criminal events that occur in this context continue to occur. Outside the primary group context, the effects of age on person offenses show themselves even more clearly. So, while we may find social conditions in which age does not have as strong an effect as usual, the isolation of such conditions does not lead to the conclusion that age effects may be accounted for by social conditions. On the contrary, it leads to the conclusion that in particular cases the age effect may be to some extent obscured by countervailing social processes.[9]

Artificial Conditions and Behavior Analogous to Crime

Theories designed to explain age effects focus on the social position of youth vis-à-vis adults, suggesting that if their situations were identical, the differences in their crime rates would disappear. One way to test such theories would be to construct an artificial environment in which age varies and the forces said to create the age relation are held constant. For example, if differential labor force participation is said to account for the age effect, we could test this thesis by creating an environment in which no one participates in the labor force. All this is more easily said than done. The closest we can come to a natural approximation of an environment that holds at least some of the putative causal variables constant is the prison. For research, prison populations have the advantage of being relatively homogeneous on many crime-causal variables, since they

[9] Our invariance hypothesis was generated by observation of the stability of the age distribution of crime over a variety of conditions. It was intended to be merely an empirical generalization stating a relation between observable variables. Defense of such generalizations in the face of variation across indicators, however minor, requires some degree of conceptualization. In the discussion here, we have substituted "tendency to commit criminal acts" for "crime," and our invariance hypothesis has thus become "the age distribution of the tendency to commit criminal acts is invariant across social and cultural conditions." This revised hypothesis is not strictly at the mercy of the facts and is therefore not necessarily contradicted by the observation that the relation between age and various *indicators* of crime is not precisely the same under all conditions. We would have to grant, for example, that removal of large portions of high-crime-rate groups from a population could reduce the crime rates for those groups (especially if [a] the individuals removed remained in the denominator of the rate or [b] those removed were selected for their higher likelihood of crime), but we would not have to grant that the data produced by such an experiment contradict or falsify our invariance thesis. Our initial inclination was to attempt to defend the age-crime relation as an empirical generalization without benefit of general concepts or theory, in order to stress the similarity of the shape of the distribution under all conditions. We have learned, however, that theory-free facts are no match for fact-free theories. Reviewers and colleagues have repeatedly demanded theory or have derived from theory statements that "contradict" our thesis. The standard way of dealing with this problem is to follow the presentation of facts with a summary of various theories. This procedure allows facts and theory to borrow truth from each other (see Galtung 1967, pp. 453–54) and is eminently satisfying. The simple fact is, however, that this solution is closed to us. We cannot simultaneously maintain that the age-crime relation is beyond current theory and suggest that this relation supports and may be derived from such theory.

are relatively homogeneous on crime.[10] As shown by figure 6, which presents prison infraction rates by age, when "practically everything" is held relatively constant, the age effect is much like the age effect in the free world.[11]

Another way to approach the problem of the confounding of age with other causal variables is to isolate an item of behavior analogous to crime but lacking at least some of its components, such as socially induced motivation. Automobile accidents satisfy some of these criteria. Such accidents do not suggest economic need or exclusion from the means of production; they do presuppose a minimum of ability, an ability that quickly reaches its maximum and then slowly but steadily declines; furthermore, the various dimensions of personality thought to be associated with accidents are not thought to be correlated with age.

Figure 7 shows the motor vehicle accident rate in New York State among those eligible to drive, by age. As is evident, these data closely parallel those for crime. We will have reason to discuss the theoretical import of these accident data subsequently.

2. THE AGE CRITIQUE OF THEORY IS UNJUSTIFIED

Most current theories of crime concentrate on the adolescent and late teen years, when the rate of crime is at or approaching its maximum level.

[10] Given that the correlations between crime-causal variables and crime are often weak, prison populations are far from homogeneous in a statistical sense. In fact, in this sense, these populations may tend to be more heterogeneous than the population as a whole. (Minorities tend to contribute more than their share to the offender population.) Nevertheless, within groups relatively homogeneous on crime, crime-causal variables tend to lose their ability to predict subsequent criminality. For example, Glueck and Glueck (1970, pp. 174–80) report that very few of the "traits and factors" that differentiated delinquents from nondelinquents in adolescence continued to differentiate offenders from nonoffenders among the delinquents followed into adulthood.

[11] Similar relations between age and rule breaking in prison have been reported repeatedly (Ellis, Grasmick, and Gilman 1974; Wolfgang 1961; Zink 1958; Mabli et al. 1979). Flanagan (1981, p. 3) reports that "the most adequately established correlate of misconduct among prison inmates is age." It is difficult to obtain age-standardized prison infraction data. The data in fig. 6 were constructed in the following fashion: the age bases represent the population of male inmates by age category in New York State prisons, as of December 31, 1975 (source: New York State 1976, table 2G). The infraction data are derived from a sample of releasees, drawn to be representative of all male releasees from New York State prisons in 1973–76. The number of releasees sampled was 758; the number of infractions recorded was 4,293. Thus, the data are limited in two respects: the infraction data reflect the experience of a sample of releasees, while the base data are for all persons confined, and the time periods are not identical. The infraction data represent codings from the institutional history file of the sampled inmates. Any recorded adjudication by either of two disciplinary bodies— the prison adjustment committee and the superintendent's proceedings—were coded. These are, essentially, all formally adjudicated infractions. Data on ages of persons in prison were available only for the catgories 16–18, 19–20, 21–29, 30–39, 40–54, and 55 and older. The midpoints of these categories were used as the basis of the figure presented in the text. As far as we can see, the biases in this procedure would tend to affect the location of the distribution rather than its form.

The general strategy is to identify or construct high- and low-rate groups, to differentiate between delinquents and nondelinquents. However differentiation is accomplished—whether by labeling, exposure to definitions favorable to delinquency, lack of legitimate opportunity, reinforcement of incipient delinquent conduct, or lack of social restraint—the result is identification of groups unusually likely to commit criminal acts.

Standard research procedure in testing such theories is to compare the actual crime rates of the groups they identify. Although in practice the theories may be difficult to test because of ambiguity or inconsistency, there is in principle little disagreement about how they should be tested. If differential opportunity is said to be the key to delinquency, one defines opportunity operationally and compares the delinquency rates of those having more with those having less. Up to the actual initiation of tests, there appears to be no necessary empirical defect in such theories. And since they are at least in principle testable, there would appear to be no necessary logical defect in them either.
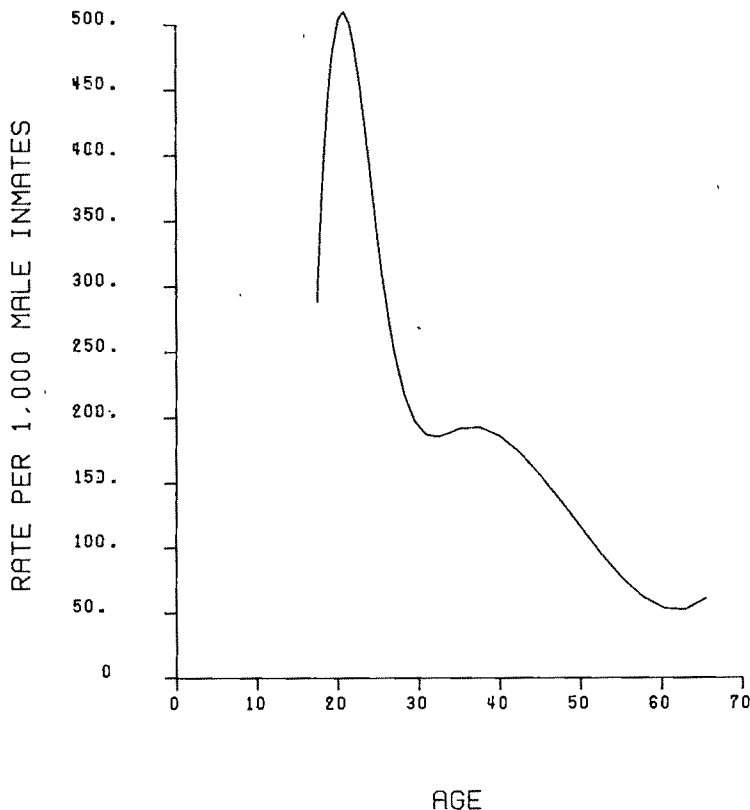
FIG. 6.—Prison infractions per 1,000 inmates, New York State, 1975. Graph presents analysis of raw data from Flanagan (1979) and New York State (1976).

Enter the brute fact, the age distribution of crime. Just at the point where the criminal group has been created, it begins to decline in size. "Maturational reform" or some equivalent unexplained process takes over. The theory is then said to be able to explain the onset of crime but unable to explain desistance from crime. Since "desistance" is equal in theoretical significance to "onset," this failing of the theory is considered to be a failing sufficiently serious to bring its explanation of the onset of crime under a cloud of suspicion: "Since most delinquents do not become adult criminals, can we assume that somehow their social bonds eventually are strengthened? How is this possible? Control theory does not adequately answer these and similar questions" (Siegel and Senna 1981, p. 139). And: "Social process theories do not account for the 'aging out' factor in delinquency. This is a fault of the . . . social structure approach as well" (Siegel and Senna 1981, p. 147).
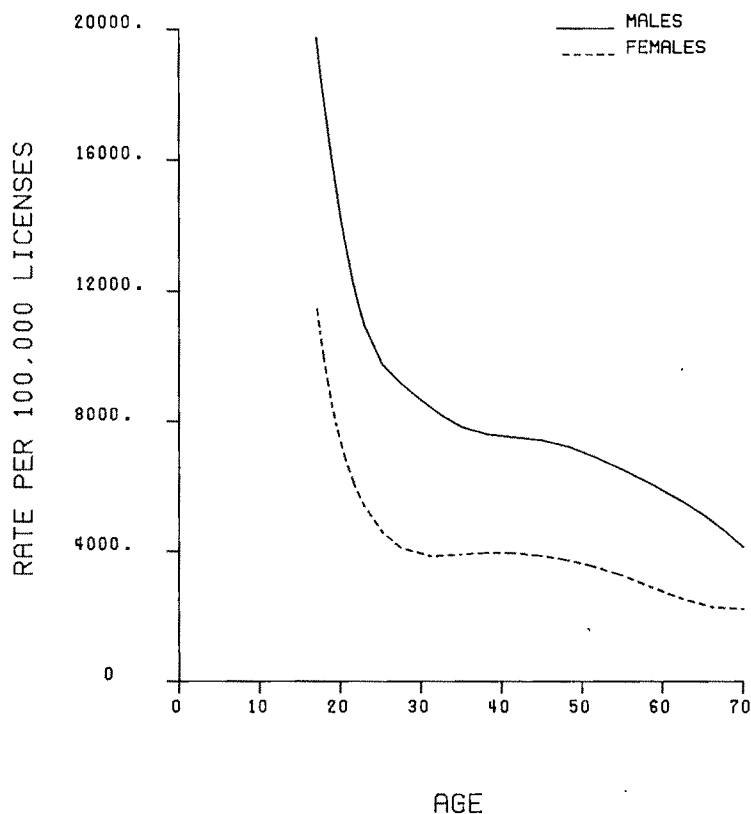
FIG. 7.—Motor vehicle accidents by age and sex. Rates per 100,000 licenses in force, New York State, 1977. (Source of data: New York State [1979].)

This by now traditional criticism should be understood for what it is: a theoretical argument dressed as a logical and empirical argument. The empirical fact of a decline in the crime rate with age is beyond dispute. The requirement that theories account for facts is also beyond controversy. But it does not follow that a theory that adequately differentiates criminals from noncriminals will also account for the effects of age. What makes the argument theoretical is that it *requires* that the age distribution of crime be accounted for by the variables explaining crime rate differences at a given time. This amounts to an assertion that the age effect on crime cannot be independent of the variables employed by an accurate theory of crime. Yet it could be that a given theory, in which the rate for the low-rate group is simply a constant proportion of that for the high-rate group, holds true at all age levels. Figure 8 illustrates this possibility. It shows a true theory unaffected by "maturational reform." This theory differentiates offenders from nonoffenders throughout the life cycle. Its failure to account for the "aging-out" factor in crime cannot therefore be taken as a "fault" of the theory since the aging-out effect occurs constantly in each group. Clearly, until evidence against this plausible hypothesis
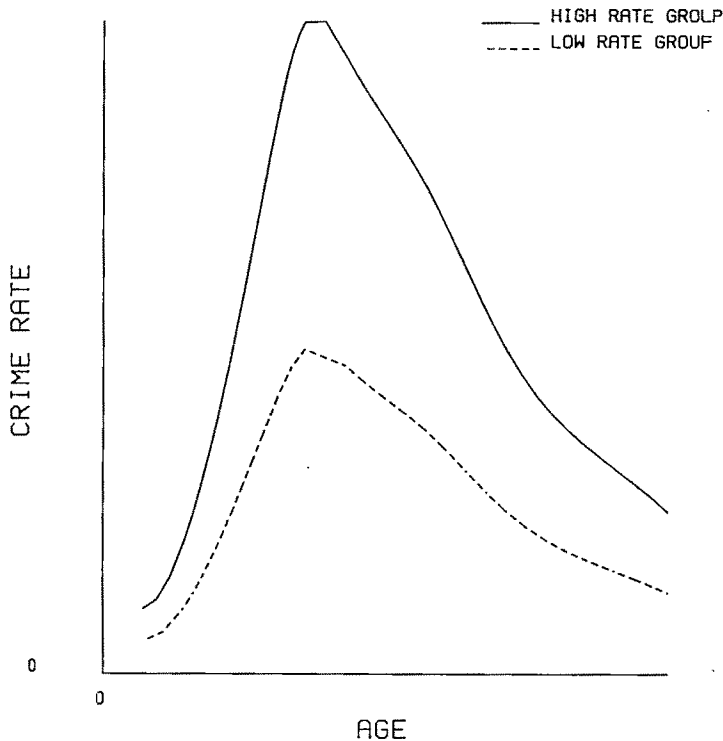
FIG. 8.—True theory unaffected by age

has been located, there is no justification for using age as a critical weapon against any current theory of crime.

This point may be illustrated by application of the logic of age-based critiques of social theories of crime to the motor vehicle accident data displayed above. No one would argue that the impact of driver training on accidents is inadequate as an explanation of variation among drivers because it fails to account for the age effect. Indeed, insurance companies that routinely give premium discounts for persons with driver training do not neglect the age variable. More generally, it is beyond question that age affects the likelihood of motor vehicle accidents regardless of the social characteristics of drivers. It should be mentioned also that the natural costs of accidents are usually far greater than the social costs and/or formal penalties. There is, therefore, no reason to believe that social control can account for the shape of the age distribution of accidents.

Thus, if the possibility depicted in figure 8 describes the actual situation, efforts to bring theories into line with the age distribution, to encompass the effects of age, will lead the theorist into assertions contrary to fact.

For example, Sutherland and Cressey (1970, p. 126) argue that the theory of differential association can account for the apparent effects of age. Presumably this means that age is correlated with exposure to particular constellations of definitions favorable to violation of law, and that in groups where there is no change in definition, there will be no change in the likelihood of crime over the life cycle. Yet research shows that, in accord with our thesis, "even with equal exposure to criminal influences, propensity toward crime tends to diminish as one grows older" (Rowe and Tittle 1977, p. 229).

Again, this fact does not invalidate the theory of differential association. On the contrary, it is exactly what we would expect were the theory (or any theory) true *and* independent of age. The reader will note that figure 8, though hypothetical, closely approximates actual subgroup differences reported elsewhere in this paper. It therefore seems safe to say that (1) the argument that theories of crime must take age into account is itself a theory of crime, and (2) the theory underlying this argument is contrary to fact.

3. THE AGE EFFECT CANNOT BE EXPLAINED WITH AVAILABLE
   CONCEPTS

Our third thesis is a corollary of the thesis that the age distribution of crime is invariant across social conditions. If the age effect cannot be even partially explained by historical trends or cross-cultural comparisons, if it is unaffected by introduction of such gross correlates of crime as sex and race, if it appears when other known causes of crime (including

crime itself) are held constant, and if it shows itself in phenomena analogous to crime that lack many of the elements typically encountered in explanations of crime, then there is reason to believe that efforts to explain the age effect with the theoretical and empirical variables currently available to criminology are doomed to failure.

The exceptional plausibility of theories of crime that focus on age is, nonetheless, hard to deny. Ironically, a major source of their plausibility appears to be the robustness or universality of the relation between age and crime. Since no one doubts that age (unlike, say, social class) is an important and powerful predictor of crime, it follows that theories capable of explaining this relation are also important and powerful (and probably true). The fact that social or cultural theories that explain the age effect automatically predict variation in the age-crime relation does nothing to diminish their plausibility. On the contrary, predicted variation serves to enhance the plausibility of the theory predicting it, whether or not such variation is actually observed. Other sources of plausibility include the ability of age theory to address historical trends as well as current differences in crime rates and its ability to use a variety of sociological perspectives in a noncompetitive fashion (different perspectives are used to account for different features of the age distribution).

Given the plausibility of age-based explanations of crime, it should be instructive to examine the logic and empirical adequacy of the most prominent theory built explicitly on age, that presented by David Greenberg (1979) in "Delinquency and the Age Structure of Society."

Greenberg first focuses on theories that address within-age variation. These theories are said to "shed little light on the relationship between crime and age" (1979, p. 589). The theories of Miller, Cohen, and Cloward and Ohlin are all argued to be defective vis-à-vis the age distribution of crime. This leaves the door open for theories that take the age distribution as the direct focus of attention. Theories said to be partially adequate in this regard are those of Bloch and Neiderhoffer (age status problems) and Matza (drift accompanied by reduction in male status anxiety). As one would expect, these partially adequate theories have difficulty explaining *within*-age variation, but they are taken by Greenberg as pointing in the right direction.

In the end, Greenberg proposes a theory combining strain and control which, he says, provides "a very plausible account of age and other systematic sources of variation in delinquent involvement . . ." (1979, p. 591). The strain comes from exclusion of youth from access to the means of production, with its resulting material deprivation and masculine status anxiety. Control enters through increasing legal penalties and increasing social integration ("stakes in conformity") with age. Greenberg focuses on four "facts" about the age distribution of crime: (1) the peak age of

crime is late adolescence, (2) the peak age declines over time,[12] (3) the peak age varies by type of offense, and (4) offenses decline in number with age. The strain component of the theory is said to explain the first three of these facts, the control element the fourth. Greenberg's argument is plausible to the extent that the cross-sectional theories on which it is based are plausible. But whatever the plausibility of the source theories, there is no empirical reason to believe that any combination of them explains the age distribution of crime. Let us turn first to the strain component in Greenberg's theory.

Adolescent Deprivation Replaces Class Deprivation

Greenberg catalogs the needs of adolescents ("clothing, cosmetics, cigarettes, alcoholic beverages, narcotics, phonograph records, transistor radios, gasoline for cars and motorcycles, tickets to films and concerts, meals in restaurants, . . . gambling" [1979, p. 593]) and argues that the deterioration of the position of adolescents in the labor force has "progressively eliminated" legitimate employment as a source of the funds required to meet these needs. The argument, then, is that adolescents are a deprived class and that their (relative) deprivation has increased steadily in recent years: ". . . teenagers [are] less and less capable of financing an increasingly costly social life whose importance is enhanced as the age segregation of society grows. Adolescent theft then occurs as a response to the disjunction between the desire to participate in social activities with peers and the absence of legitimate sources of funds needed to finance this participation" (1979, p. 594).

The plausibility of "poverty" explanations of crime is beyond question (see, e.g., Blau and Blau 1982). Whether applied to differences among adolescents or between adolescents and adults, these theories make a great deal of sense. We should note, however, that there is no more evidence for Greenberg's theory than for the general strain or motivational theories from which it is derived. These general theories have consistently shown themselves to be less than adequate in dealing with basic cross-sectional facts about crime (Kornhauser 1978). For example, they automatically predict that employed teenagers will be less likely than unemployed teenagers to steal, that teenagers with access to legitimate funds will be less likely to steal, and so on. The data from delinquency research simply do not support these predictions (West and Farrington 1977; Hir-

[12] Daniel Glaser (personal communication, 1982) predicts an increase in the peak age and a flatter age distribution over time. Both of these predictions, he notes, "contradict our theme of a constant age-crime relationship." Both of Glaser's predictions also contradict those advanced by Greenberg. As we have noted, neither set of predictions is consistent with the evidence. Therefore neither "contradicts" our theme.

schi 1969, pp. 188–89). In fact, they show a relation between individual affluence and delinquency opposite to that predicted by strain-poverty-deprivation theory.

If a theory mispredicts cross-sectional differences, there is little reason to think that it is adequate as an explanation of age differences in crime. But there is direct evidence on the age aspects of the theory. Tittle (1980, p. 92) shows that the age distribution of self-reported offenses is unaffected by controls for sex, race, marital status, socioeconomic status, state of residence, size of place, religion, place of childhood residence, family background, or, most important in the present context, labor force status. In none of the 36 demographic categories identified in Tittle's analysis does the negative correlation between age and deviance drop below −.40 (gamma) (1980, p. 92). (Critical to Greenberg's theory is the fact that two of Tittle's items measure theft.)

A complex strain theory is hard to test. Rarely, in fact, do such theories restrict their attention to combinations of variables currently available in the research literature. For this reason, even the rare data analyzed by Tittle (and by Rowe and Tittle 1977) are not sufficient to the diverse claims of Greenberg's presentation, a major feature of which is that the age distribution of crime in modern capitalist society differs from the age distribution of crime in earlier periods, or, by extension, in less industrialized, less capitalistic societies.

But the similarity between the age-crime distributions through time (1835–1980) and across place (Argentina, the United States, France, Sweden, Japan, England, and Wales) is remarkable (see figs. 1–3 and n. 5 above). In shape or form, they are virtually identical. The major or only difference is in the location of the curves along the age axis. In early 19th-century England and France, crime peaked at a later age than it does in late 20-century America. What is the significance of this fact?

Variation in the age of maximum criminality has been widely noted (e.g., Sutherland and Cressey 1970, p. 122). The peak age varies across time and place and by type of offense. Person offenses tend to peak later than property offenses and have done so for some time (Quetelet 1969, p. 93).

Interestingly enough, Greenberg begins his discussion of the *crime-type* difference in peak age with what may be seen as an explanation of the *over-time* difference: "Over time, the 'democratization' of the family has reduced the age at which given levels of autonomy are acquired" (1979, p. 596). But the theoretical problem is to explain the later peak age of person offenses—rape, robbery, homicide, assault—and the generally slower decline of such offenses after the peak age has been attained: the answer Greenberg provides is "masculine status anxiety" engendered by "the contradiction between the school's expectations of docility and submission

569

to authority, and more widely communicated social expectations of masculinity" on the one side and "inability to fulfill traditional sex role expectations" (1979, pp. 604–5) on the other: "One would expect masculine status anxiety to appear with greatest intensity and to decline most slowly in those segments of the population in which adult male unemployment is exceptionally high. This conforms to the general pattern for violence offenses . . ." (1979, p. 605). There are difficulties with this explanation, too. As we argued above, if the self-report data (Tittle 1980; Tittle and Rowe 1977) showing no difference in peak age for property and person offenses are taken at face value, the difference in the peak years for person and property offenses in official data is a function of differential response by age rather than a function of differential causal factors operating to produce person and property offenses. Evidence of differential causal factors by type of crime is nonexistent. Indeed, the persistent search by criminologists for "types of offenders" has met with repeated empirical failure (see, e.g., Wolfgang et al. 1972; Peterson and Braiker 1980). Since the data strongly imply that the same individuals are involved in both "types" of offenses, explanations such as Greenberg's that posit a shift in the operative causal factors with age confront empirical (as well as logical) difficulties.

Social Control Explains the Decline in Crime with Age

Having cataloged the various motivational factors in crime that may reach a peak at the point of maximum criminality, Greenberg turns to a discussion of the cost or social control aspects of criminality. He begins by rejecting the notion that internal controls or moral inhibitions increase with age sufficiently to account for the age-related decline in delinquent conduct. External costs are, however, another matter: "Parents and teachers are generally willing to write off a certain amount of misbehavior as 'childish mischief,' while enormous caseloads have forced juvenile courts in large cities to adopt a policy that comes very close to . . . 'radical nonintervention' for all but the most serious cases. . . . [However], as teenagers get older, the potential costs of apprehension increase; victims may be more prone to file a complaint, and police to make an arrest. Juvenile court judges are more likely to take a serious view of an older offender . . ." (1979, pp. 606–7). Nor is the justice system the only source of increased costs with increasing age: "Just as the costs of crime are escalating, new opportunities in the form of jobs, marriage, or enlistment in the armed forces create stakes in conformity and . . . may also relieve problems of masculine status anxiety" (1979, p. 607). (Differences in the age distribution between blacks and whites, lower and middle class, and

by type of offense are all said to be consistent with this thesis [1980, pp. 607–8].)

There are several difficulties with this argument about the effects of external social control. (1) No data on the age distribution of crime show the discontinuous decline in the rate of crime one would expect were formal sanctions responsible for the decline. Since penalties for crime depend on age, one would expect to observe in the data the transition of people from one set of sanctions to another. But the data do not reflect such movement. Instead, they evidence a continuous, monotonic decline in crime rates with age once the peak has been attained (whatever this peak age may be). In fact, Rowe and Tittle (1977, p. 231) show that all four types of deviance they investigated (theft, gambling, assault, tax cheating) are negatively correlated with age when delinquency of acquaintances in childhood, social integration, utility (of the behavior), moral commitment, and sanction fear are partialled out. They conclude that their results "clearly show that the explanations offered, when each is considered alone, are inadequate to account for the persistent age/crime relationships" (1977, p. 232). (2) The factors Greenberg adduces to explain desistance offer a plausible account of crime rate differences between late adolescence and early adulthood, between, say, 19 and 24, but they do not provide a plausible account of the similar decline in crime rate between, say, 29 and 34. Yet in all of the data, including the data provided by Greenberg, the decline in this five-year period is as steep as in the earlier five-year period. (3) Greenberg's account of the situation of juveniles vis-à-vis the criminal justice system clearly suggests that the true crime rates in early adolescence may be higher than those in late adolescence, that, in other words, the peak age of "criminality" may be even earlier than crime statistics suggest. If so, the economic need, the masculine status anxiety, and the formal-external control hypotheses are called into question. Fourteen-year-olds do not need money for cigarettes, alcoholic beverages, narcotics, gasoline for their cars, and for gambling—unless, as Greenberg's description of their "needs" often suggests, they are already delinquent—or if they do need such things, it is clear that they are likely to need them less than, say, 17-year-olds. For, the fact is, the use of cigarettes, alcohol, narcotics, and gasoline increases steadily throughout the teen years, while the rate of clearly criminal offenses (especially, in official data, theft offenses) peaks in middle adolescence and then begins to decline.

The best that can be said is that contemporary arrest data peak a few years earlier than conviction or incarceration data from earlier periods. For purposes of discussion, let us assume that the maximum age of criminality has indeed declined slightly in the past 60 years. (Contemporary

571

incarceration data may even cast this assumption in doubt.) The assumption of a declining age of maximum criminality is taken to reflect change in the social position of youth. The social position of those at the current maximum is not identical to the social position of youth at the same age some time in the past. Therefore, it seems to follow that the relation of age to crime has changed. But what if we compare the social position of youth at the current maximum with youth at the historical maximum? What do we conclude? Apparently, the social situation of the two groups is similar—otherwise, each time period would require a unique explanation of the period of maximum criminality (and any age group could, in principle, occupy this position).

Thus, an explanation that argues that high rates among contemporary youth are a product of exclusion from legitimate means of satisfying peer and media engendered needs (Greenberg 1979) would have to argue that 25-year-olds in Wales in 1842 were also so excluded and were also so vulnerable to peer and media consumption pressures. Thus, even if a nontrivial difference in peak age were present in the data, an explanation of the current pattern that did not recognize the existence of a virtually identical form to the age distribution in earlier periods would encounter serious logical and empirical difficulties. It is much simpler and more consistent with the data merely to assume that the age effect is virtually invariant over the range of social conditions for which data are available, and that small changes in the peak age may reflect either (1) the earlier emancipation or (2) the earlier physical development of youth. That is, with the exception of minor fluctuation in mode, the essential feature of the age distribution is extraordinary stability.[13]

In sum, Greenberg's argument concerning the theoretical importance of age is explicit and rests on two principles:

1) age variation may help to test delinquency theories constructed to explain other sources of variation, such as class or sex. Since these other sources of variation can be explained in many ways, the adequacy with which

[13] Whether the peak age is 17 or 19, or whether this peak age varies somewhat by offense or by sex, may be quite insignificant in contrast to the stability of the major parameters of the age-crime distribution. In other words, variation in the location of the curve on the age axis across time, place, and such demographic factors as sex or even type of offense may say little or nothing about the impact of age on criminality. If age affects criminality within all of the groups so identified (and available evidence indicates that it does), then it is, to say the least, unusual scientific logic to conclude that it therefore has no effect on criminality. Most textbook discussions of the age distribution in a theoretical context rely on assumed differences for demographic subgroups. Theoretical discussions of the supposed interaction of sex and age on crime are good examples. Textbooks perhaps typically compare "rates of increase" for boys and girls on particular offenses, thus suggesting considerable flexibility in the age distribution by sex. Elaborate explanations of the changing structure of society, or of role anxiety, are then constructed to account for these fluctuations. All of which tends to obscure the basic fact that the age-crime relation does not vary by sex.

rival theories explain age variation may help us to distinguish among them . . . [and]

2) any explanation of age variation in criminality based on psychological reactions to physiological changes accompanying adolescence would be difficult to reconcile with the great variation in delinquency involvement among juveniles as well as the lateness of peak involvement in violence offenses. If age is relevant to criminality, the link should lie primarily in its social significance. [1979, pp. 588–89]

Thus, Greenberg explicitly denies the relevance of nonsocial explanations of the age distribution, relying mainly on variation in criminality within age groups as evidence against such explanations. However, if, as we have previously suggested, it is illegitimate to use age variation against social theories, it is equally illegitimate to use within-age variation against nonsocial theories. We must then disagree with Greenberg on both counts. No rule of logic requires that explanations of one correlate of crime also be explanations of other correlates. If the effect of age on crime does not interact with other effects, explanations of the age effect may not explain the effects of other variables. Once again, then, we see delinquency theorists dressing theoretical arguments as logical and/or empirical arguments, a most distressing habit, since it implies that alternative theories are clearly defective on both logical and empirical grounds when they may be defective on neither.

The robustness of the relation of age and crime—across time, place, and social condition—grants a high degree of plausibility to explanations of crime, such as Greenberg's, that focus on age. Ironically, the very fact that gives such theories their plausibility also falsifies them. A ubiquitous relation falsifies explanations the moment they are advanced, and the ubiquity of the age relation to crime is phenomenal.

4. AGE HAS A DIRECT EFFECT ON CRIME

A direct-effect hypothesis makes sense only in the context of a restricted set of competitive or intervening variables. We would not argue that no mechanism can be found to account for the effects of age; we argue only that no such mechanism is to be found in current criminological research or theory. Since this argument follows directly from what has been said before and is based on data previously discussed, further elaboration seems unnecessary.

5. CONCEPTUALIZATION OF THE AGE EFFECT IS LARGELY REDUNDANT OR MISLEADING

An increasingly prominent focus of contemporary criminology is the career criminal. Statistics to the effect that a small percentage of offenders ac-

count for a large percentage of offenses are routinely cited in support of this concept. Taken at face value, the career criminal notion suggests division of the criminal population into discrete categories: those who offend occasionally or sporadically for a typically brief period of time and those who offend regularly over an extended period of time. If the division proves valid, two general age-crime distributions could be extracted from the data. For the first group, the age distribution might well approximate the now familiar positively skewed, unimodal distribution in evidence throughout this paper. For the second group, the career criminals, several distributions are possible, depending on how such careers proceed, and several such distributions have in fact been postulated. The most common asserts that career length is a function of "age of onset"—the younger the age of onset, the older the age of desistance. "Boys first convicted at the earliest ages tended to become the most persistent offenders as adults" (Farrington 1979, p. 12; see also Shannon 1978; Department of Justice 1981–82). (The underlying model here is that the more serious or persistent the offender, the earlier in life will this fact be evident.) An alternative conception suggests a fixed length to the criminal career, so that those who start early will also finish early. This suggests that all careers are of approximately equal length. This distribution is often associated with the concept of "burnout." Other models are possible within these generic types. For present purposes, the question is whether the age distribution of crime can shed light on this way of looking at the phenomenon.

Let us focus directly on the major elements of career descriptions: "age of onset," "age of desistance," and their derivative, "length of criminal career."

Age of Onset

The age-of-onset terminology appeared early in the criminological literature (see Goring 1913; Wootton 1959, chap. 5). As originally conceived, the concept involved the isolation of a group of offenders and subsequent identification of the age at which their criminality first appeared. The logic for investigating the etiological significance of age on the basis of the age-of-onset notion was borrowed from etiological studies of disease:

> . . . as a subject for independent inquiry, the present age of any sample of criminals appears to have no more direct and special statistical import, in relation to criminological questions, than the present ages of a sample of tubercular subjects would have direct statistical bearing upon problems of tubercular disease. . . . And just as in the one case a solution of this etiological problem depends upon the analysis of statistics relating, not to the ages of subjects during the course of disease, but to their ages at the time of its onset, so, to elucidate the etiological relation of age to crime,

> the statistics we require are not the ages of criminals at any period of their
> career, but statistics of the age distribution of criminals at the time of their
> first offense. [Goring 1913, p. 201]

When research is based on a sample of continuing offenders and retro-
spective procedures, onset notions make sense. However, Goring himself
used an alternate procedure to study age of onset. Although this alternate
procedure is repeatedly used in contemporary research, it is inconsistent
with the logic of the original idea: Goring distinguished habitual from
first-time offenders and compared these groups with respect to age of
onset. When the research design was altered to avoid exclusive concen-
tration on persistent offenders and retrospective accounts of their "ca-
reers," the terminology and logic of onset were retained. But when samples
include those who persist in offending and those who do not, age of onset
becomes a problematic idea, and the effects of age-of-onset differences
are easily confused with differences in rates of offending.

For example, a common mistake in traditional thinking about age of
onset has been to confuse the finding that a given group is more likely to
produce offenders with assertions about the characteristics of its offenders.
Thus, if a larger proportion of blacks than whites commit offenses at an
early age, this is read as meaning that blacks begin offending at an earlier
age than whites. The question then becomes, Why do blacks start earlier
than whites? If blacks start earlier than whites, there is, given career
notions, reason to believe that blacks may be more serious offenders than
whites and that, at any given age, they will be "further along" in their
careers than white offenders. This logic then leads to comparisons of
black offenders with white offenders when, in fact, there may be no
evidence of differences between the two that require explanation.

Figure 5 shows delinquency rates for white and nonwhite males in the
Wolfgang et al. (1972) birth cohort. Note that in these data the higher
rate for nonwhites at early ages persists throughout the age range available
in the sample, and that there is no difference in the form of the age
distributions of delinquency for nonwhites and whites. Note too that
higher rates for nonwhites at ages 10–12 may be taken as evidence that
nonwhite delinquents "start earlier" than whites. Wolfgang and his col-
leagues address this question directly:

> Nonwhites generally incur their first police contact at an earlier age than
> whites. The respective mean ages of onset are 13.4 for nonwhites and 14.3
> for whites. Moreover, the percentage of nonwhite youth arrested for the
> first time is higher than the corresponding percentage of white youth at
> each age of onset category between 7 and 13 years, whereas a higher
> proportion of white than nonwhite youth fall into the 14–17 age categories.
> Thus, up to and including age 13, 48.7 percent of the nonwhite youth,
> compared to 30.8 percent of the white youth, have their initial police con-

tact. The trend is reversed between 14 and 17 years. [1972, p. 135; see also U.S. Department of Justice (1981–82), which reports similar findings from subsequent cohorts studied by Wolfgang]

Figure 9 compares prevalence rates for nonwhites and whites calculated from the Wolfgang et al. data. These rates appear to be perfectly consistent with the age distributions of criminality presented in figure 5. The proportion of nonwhites "beginning" delinquency, like the proportion committing delinquent acts, is greater than the proportion of whites at all age levels. If we calculate a measure of association between race and onset at each age from seven to 17, the results are as follows: .73, .62, .68, .67, .67, .67, .64, .59, .54, .46, .43.[14] These results are consistent with the Wolfgang et al. discussion just quoted. They show that when delinquency is measured by the first offense, something happens to the
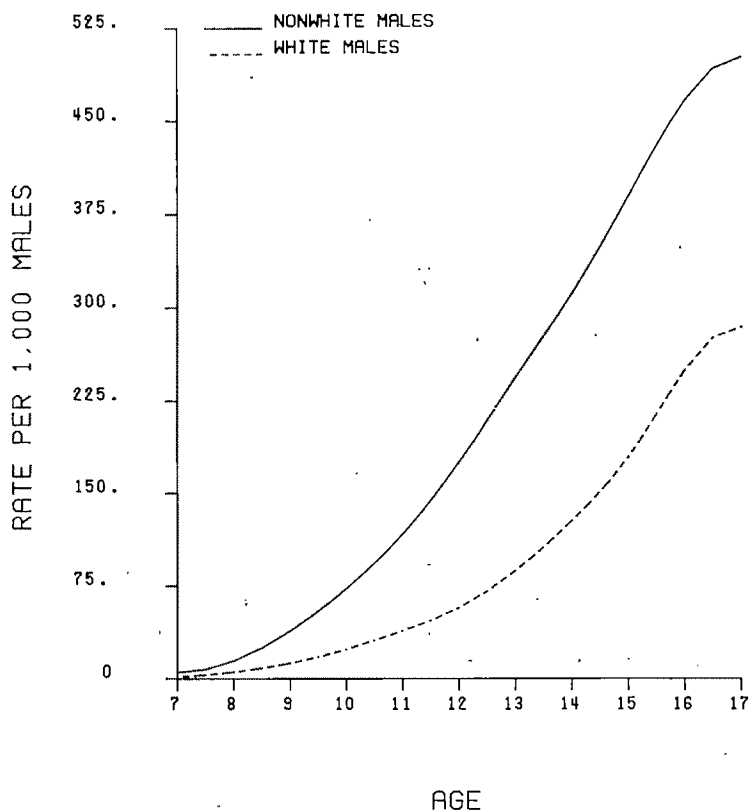
FIG. 9.—Prevalence rates by age and race. (Source of data: Wolfgang et al. [1972, p. 90].)

[14] The measure of association is NW − W/NW, where NW and W are the nonwhite and white delinquency rates.

race/"delinquency" relation at about age 13. The difference between non-whites and whites begins to decline. Why? As figures 5 and 9 and the measures of association show, nonwhites are progressively more likely than whites to be and to become delinquent throughout the age range covered (at least to age 16). At about age 13, however, the proportion of nonwhites who have already become delinquent becomes sufficiently large that the relative increase in nonwhite first offenders cannot keep pace with the relative increase in white first offenders. For nonwhites to keep pace with whites between ages 13 and 17, about 55% of the entire non-white population (almost three-quarters of those eligible) would have to commit a *first* offense during this period. The resulting difference in relative increase, a logical necessity given differences in rates of delin-quency, produces all of the age-of-onset differences described by Wolfgang and his colleagues. (In fact, at age 12 and before, the mean age of onset for nonwhites and whites is identical!)

The fact is that the longitudinal design with its age-of-onset terminology unduly complicates our explanation of its findings. In (relatively inex-pensive) cross-sectional terminology, this major finding of longitudinal research may be expressed as follows: at all ages, nonwhites have higher rates of crime than whites. At any given age, therefore, nonwhite of-fenders are more likely than white offenders to have committed prior offenses (to be career criminals?). Since the proportion of nonwhite of-fenders with prior offenses approaches the theoretical limit faster than the same proportion for white offenders (because of the higher crime rate for nonwhites), at some point the proportion of "new" offenders among white offenders will be larger than the proportion of "new" offenders among nonwhite offenders. This will suggest to the unwary that whites "start later" than nonwhites. If whites start later than nonwhites, it fol-lows that nonwhites start earlier than whites. But we have already es-tablished that the latter statement is in no meaningful sense true, just as the former statement, being a statistical necessity, is in no theoretical sense useful.

In short, it seems clear that age-of-onset terminology can be highly misleading; it can obscure otherwise straightforward facts and produce in their stead "findings" that defy rational explication. In the present case, it suggests differences between nonwhite and white offenders that are not supported by the data. As with all "career" notions, the age-of-onset question seems to lead to comparisons of offenders across groups, when the proper comparisons include those who do not offend as well.[15]

[15] A similar conclusion has been reached by those studying violence prediction. For example, Monahan argues (1981, p. 111) that "the relevance of race in a person with an extensive record of violence appears minimal or nonexistent. Whatever their race, people with such records have a higher probability of future violent behavior. Such findings lead one to 'emphasize the unimportance of race as a determinant [of future violence] once the individual has been identified as a delinquent' (Hamparian, Schuster, Dinitz, and Conrad, 1978:133)."

Age of Desistance

Implicit in the concept of career is some notion of behavioral stability. In most careers, end points tend to be fixed by convention or by the physical demands of the profession. Since conceptions of the natural length of a criminal career are hard to come by, the approach to this question in criminology has tended to be radically empirical. A career lasts as long as the offender continues to offend. When, then, do offenders stop offending? One way to answer this question is to follow a group of offenders over time, recording their criminal offenses. When a set period of time elapses without an offense having been committed, the offender is said to have abandoned his career. Age at last offense marks the end point of a career in crime. When age of desistance is charted, it turns out that there is considerable variation. What accounts for the fact that some offenders quit earlier than others?

This kind of question is assumed to be of singular significance and to be different from the kinds of questions criminologists uninterested in careers are likely to ask. It therefore requires a peculiar set of independent variables. Historically, the major variable in this set has been age of onset. (Age of onset serves two functions: it defines and it may explain the length of a criminal career. Separation of these two functions is often difficult.) Consider the following: "A central issue for criminal careers research is the identification of factors that discriminate between people who do and people who do not continue criminal activity after their initial police contacts. . . . Studies have found the characteristics of juvenile criminality to be the most reliable predictor of an adult criminal career. Those who engage in serious crime at an early age are the most likely to continue to commit crimes as adults" (Petersilia 1980, pp. 346–47). In the career criminal literature, this empirical relation is easily translated into a pressing theoretical issue: "But what explains an early onset of juvenile crime? Is it peer and family relationships?" (Petersilia 1980, p. 374).

Given what we know about age of onset, what would we predict about its relation to age of desistance? Recall that high-rate groups have high rates at all ages. From this it follows that high-rate groups will have high rates of recidivism (repeated offending). Since repeated offending defines a criminal career, it follows that high-rate groups will have high rates of "career offenders." It follows further that there is nothing of particular etiological significance in the concepts of "age of onset," "age of desistance," and "criminal career"—that there is nothing in these concepts not found in the concept of criminality itself.

A possible empirical objection to our conclusion would be a finding that rates of desistance are a function of rates of initial offending such that those who "start earlier" "quit earlier." Invariant age distributions

preclude this possibility. No group with relatively high rates early and relatively low rates late has been discovered.

6. THE LONGITUDINAL STUDY IS NOT REQUIRED TO STUDY THE
CAUSES OF CRIME

Age, career, and associated notions are considered so important by contemporary criminologists that the longitudinal or cohort study is approaching required status. Longitudinal logic is sometimes said to be necessary for adequate theory (Elliott et al. 1979), and longitudinal research is often said to be "superior to cross-sectional if one is primarily interested in drawing causal inferences" (Petersilia 1980, p. 337).

If our position is correct, the emphasis on this research design is unjustified and potentially misleading. We readily grant that the bulk of the evidence against exclusive reliance on longitudinal studies has been provided by longitudinal studies themselves. These studies repeatedly show the value of cross-sectional research and the risks of longitudinal analysis. For example, Shannon (1978) analyzes and compares 1942 and 1949 cohorts in Racine, Wisconsin. The purpose is to determine "which categories of people: are most likely to engage in delinquent behavior, will cease delinquent behavior as they grow older, or will continue into adult criminal activity" (1978, p. 3). We have argued that the first question leaves nothing for the others. Shannon appears to disagree, asserting that "age at first police contact is the best predictor of juvenile delinquency seriousness scores, these scores in turn the best predictor of intermediate seriousness scores, and finally, intermediate scores the best predictor of adult seriousness scores" (1978, p. 5). He also says that historical differences account for an "earlier onset of careers" (1978, p. 7) in the 1949 cohort and that the existence of "a relationship between more frequent and more serious [police] contacts early in life and continuity in careers cannot be denied" (1978, p. 11). Let us examine these statements within our perspective. "Age at first police contact" must mean "police contact before a given age" (otherwise Shannon would be attempting to predict juvenile delinquency seriousness scores for offenders whose first contacts were in adulthood). This statement thus reduces to the assertion that delinquency predicts delinquency. The "earlier onset of careers" in the 1949 cohort simply reflects a higher rate of crime in this cohort (and tells us the source of the idea that delinquents today "start earlier" than they once did). The relationship between frequent and serious police contacts early in life and "continuity in careers" can be similarly explained.

The correlates of "police contact at an early age" identified by Shannon in his longitudinal study (1978, p. 5) are consistent with the correlates of delinquency and criminality repeatedly reported in cross-sectional re-

search, whether or not such research had access to age-of-onset information. This is true because "delinquency at an early age" is nothing more than delinquency, and age does not interact with any known causal variables in its effect on crime. Therefore, identification of the causes of crime at one age suffices to identify them at other ages as well, and little substantive benefit accrues to the longitudinal design for the study of crime causation.

Life-Course Explanations

Age is correlated with important events thought to be related to crime, such as leaving school, marriage, and gainful employment, but its effects on crime do not appear to depend on these events. Age affects crime whether or not these events occur. We described the data on employment status earlier. Good research indicates that marriage does not affect delinquency either:

> Marriage has often been invoked as the reason for the observed decrease in convictions after age 18, and indeed as the most effective treatment for delinquency. The Cambridge study found that both official and self-reported delinquency decreased between 18 and 21. Men who married during this period were compared with those who stayed single, to see if the married group decreased more. The groups didn't differ in official or self-reported delinquency at age 21, even after attempts were made to match them up to the date of the marriage. [Farrington 1979, p. 314; see also Tittle 1980]

Although not designed as direct tests of life-course questions, studies of crime during military service are, in our view, also consistent with the argument that life-course change cannot account for the age effect (Glueck and Glueck 1968, chap. 12). The persistence of the age effect in incarcerated populations casts doubt on the assumption that such status changes as marriage, parenthood, or employment are responsible for decreases in criminality associated with age. Perhaps more fundamentally, the stability of the age effect across societies and demographic groups would not be expected were life-course factors responsible for an "apparent" age effect.

Theories that try to explain the age effect by relying on life-course events will always sound plausible. Their plausibility stems from the fact that the age effect is confounded with the effects of its correlates. (For example, marriage and "settling down" do go together because age predicts them both.) Age is correlated with beliefs and practices themselves correlated with crime—for example, respect for authority, punitiveness toward offenders, church attendance—but we believe that these correlates are not responsible for the age effect. Although crime-relevant beliefs and practices indeed vary greatly over the life cycle, the data suggest the

effects of age will be found in all categories of these beliefs and practices. Once again, the plausibility of explanations of the age effect based on such correlates results from the universal tendency to assign the effects of age to its correlates. The statistical difficulties inherent in this tendency are obvious once it is realized that none of these correlates can compete with age in predicting criminality.

Implications

Age is everywhere correlated with crime. Its effects on crime do not depend on other demographic correlates of crime. Therefore it cannot be explained by these correlates and can be explained without reference to them. Indeed, it must be explained without reference to them.

Although correlated with crime, age is not useful in predicting involvement in crime over the life cycle of offenders. For predicting subsequent involvement, to know that a child of 10 has committed a delinquent act is no more useful than to know that a child of 15 has done so. The implications of this fact for contemporary research practice are profound. It denies, for example, the suggestion (at the heart of the longitudinal survey and the career criminal notion) that "prevention and treatment efforts should be concentrated on those boys who begin their criminal careers early in life" (Farrington 1979, p. 301).[16]

Our argument also implies that the traditional division of the etiological problem into juvenile and adult segments is unlikely to be useful. Because

[16] We are grateful to Alfred Blumstein for stubbornly arguing this point until we were at last convinced that he was correct. Research evidence apparently contrary to this position is frequently reported. For example, "based on a preliminary analysis of the subjects in the 1959–1965 cohort who have already reached adult status, the researchers [Wolfgang et al. 1972] find that the age of onset of criminal activity is a key predictor of the offender's subsequent career as an adult. . . . The most active offenders in the Philadelphia sample commit their first offense around the ages of 10 or 11, and the least active commit their first—and generally—only offense at age 17" (U.S. Department of Justice 1981–82, p. 11). Let us try once more to clarify this issue. We say age is irrelevant in predicting subsequent criminality. Longitudinal researchers appear to say age is the key predictor. They present evidence. What does this evidence look like? Once again, the evidence is a relation between delinquency as measured at various points in time, say t-1, t-2, and t-3. If early onset is crucial, does this mean that t-1 delinquency predicts t-3 delinquency better than does t-2 delinquency? The answer, consistent with common sense and the results of longitudinal research, is obviously no. Is an offender who commits five and only five offenses at age 17 less likely to commit offenses as an adult than an offender who commits five and only five offenses at age 12? Again, the obvious answer is no. What, then, does it mean to say that age of onset is a key predictor of adult criminality? It probably means that *looking back* over the careers of offenders, one finds that those who have committed many offenses over a long period of time have also committed offenses when they were quite young. If so, it is (1) not clear how things could be otherwise and (2) not clear why longitudinal studies are required to look back at the records of offenders. Coupling onset and career notions with longitudinal research clearly suggests the possibility of identifying career offenders at the onset of their careers. As far as we can determine, no longitudinal study to date has identified, or even attempted to identify, the onset of a career in crime at the time of onset.

the causes of crime are likely to be the same at any age, the choice of sample should depend on the complexity of the theoretical argument and the causal analysis it presupposes. Resources should not be devoted to establishing the effects of a variable whose influence on crime is noncontroversial and theoretically uninteresting, especially when, almost by definition, examination of the effects of this variable precludes adequate examination of the effects of theoretically intriguing variables.

Funding agencies seem convinced by researchers that the longitudinal study is necessary for the proper study of crime (see, e.g., National Institute of Law Enforcement and Criminal Justice 1979). The alleged necessity of the longitudinal study is apparently based on a combination of substantive and methodological considerations. The major substantive consideration appears to be the age effect. The methodological considerations derive from the experimental model, because of which it is claimed that the longitudinal design is unique in its ability to resolve the question of causation. Our critique here focuses on the substantive justification for the longitudinal study. At the same time, however, we are not convinced that the longitudinal study offers solutions to causal questions commensurate with its costs. As we have shown, the conceptual apparatus generated by longitudinal thinking has been very misleading. This design has been oversold to criminology at high substantive and economic costs.

REFERENCES

Allen, Harry E., Paul Friday, Julian Roebuck, and Edward Sagarin. 1981. *Crime and Punishment*. New York: Free Press.
Besson, Antonin. 1961. *Seuils d'âge et législation pénale*. Paris: University of Paris.
Best, Harry. 1930. *Crime and the Criminal Law in the United States*. New York: Macmillan.
Blau, Judith R., and Peter M. Blau. 1982. "Metropolitan Structure and Violent Crime." *American Sociological Review* 47:114–29.
Blumstein, Alfred, and Jacqueline Cohen. 1979. "Estimation of Individual Crime Rates from Arrest Records." *Journal of Criminal Law and Criminology* 70:4.
Callcott, Mary Stevenson. 1935. *Russian Justice*. New York: Macmillan.
Connor, Walter. 1970. "Juvenile Delinquency in the U.S.S.R.: Some Quantitative and Qualitative Indicators." *American Sociological Review* 35:283–97.
DeFleur, Lois B. 1970. *Delinquency in Argentina*. Pullman: Washington State University Press.
Elliott, Delbert S., Suzanne S. Ageton, and Rachelle J. Canter. 1979. "An Integrated Theoretical Perspective on Delinquent Behavior." *Journal of Research in Crime and Delinquency* 16:3–27.
Elliott, Delbert S., Suzanne S. Ageton, and David Huizinga. 1978. "1977 Self-reported Delinquency Estimates by Sex, Race, Class, and Age." Mimeographed. Boulder, Colo.: Behavioral Research Institute.
Ellis, Desmond, Harold G. Grasmick, and Bernard Gilman. 1974. "Violence in Prisons: A Sociological Analysis." *American Journal of Sociology* 80:16–43.
Empey, LaMar T. 1978. *American Delinquency*. Homewood, Ill.: Dorsey.
Farrington, David. 1979. "Longitudinal Research on Crime and Delinquency." *Crime and Justice: An Annual Review of Research*, edited by Norval Morris and Michael Tonry. Vol. 1. Chicago: University of Chicago Press.

Flanagan, Timothy. 1979. "Long-Term Prisoners." Ph.D. dissertation, State University of New York at Albany.

———. 1981. "Correlates of Institutional Misconduct among State Prisoners." Mimeographed. Albany, N.Y.: Criminal Justice Research Center.

Galtung, Johan. 1967. *Theory and Methods of Social Research*. New York: Columbia University Press.

Glaser, Daniel. 1978. *Crime in Our Changing Society*. New York: Holt, Rinehart & Winston.

Glueck, Sheldon,.and Eleanor Glueck. 1968. *Delinquents and Nondelinquents in Perspective*. Cambridge, Mass.: Harvard University Press.

———. 1970. *Toward a Typology of Juvenile Offenders*. New York: Grune & Stratton.

Goring, Charles. 1913. *The English Convict*. Montclair, N.J.: Patterson Smith.

Greenberg, David F. 1979. "Delinquency and the Age Structure of Society." Pp. 586–620 in *Criminology Review Yearbook*, edited by Sheldon L. Messinger and Egon Bittner. Beverly Hills, Calif.: Sage.

Hirschi, Travis. 1969. *Causes of Delinquency*. Berkeley and Los Angeles: University of California Press.

Jacquart, Camille. n.d. *La Criminalité belge: Essais de statistique morale, 1863–1909*. n.p.

Jensen, Gary F., and Dean G. Rojek. 1980. *Delinquency: A Sociological View*. Lexington, Mass.: Heath.

Kornhauser, Ruth. 1978. *Social Sources of Delinquency*. Chicago: University of Chicago Press.

Mabli, Jerome, Charles Holley, Judy Patrick, and Justina Walls. 1979. "Age and Prison Violence." *Criminal Justice and Behavior* 6:175–86.

McClintock, Frederick H., and H. Howard Avison. 1968. *Crime in England and Wales*. London: Heinemann.

McCord, Joan. 1979. "Some Child-rearing Antecedents of Criminal Behavior in Adult Men." *Journal of Personality and Social Psychology* 37:1477–86.

Mannheim, Hermann. 1965. *Comparative Criminology*, vol. 2. London: Routledge & Kegan Paul.

Matza, David. 1964. *Delinquency and Drift*. New York: Wiley.

Monahan, John. 1981. *Predicting Violent Behavior*. Beverly Hills, Calif.: Sage.

National Institute of Law Enforcement and Criminal Justice. 1979. "Criminal Justice Research Solicitation: Interdisciplinary Research on the Correlates, Causes and Control of Criminal Violence." Mimeographed. Washington, D.C.: Department of Justice.

Neison, Francis G. P. 1857. *Contributions to Vital Statistics*. London: Simpkin, Marshall.

New York State. 1976. *Characteristics of Inmates under Custody*. Albany, N.Y.: Department of Correctional Services, Division of Program Planning, Evaluation, and Research.

———. 1979. *New York State Statistical Yearbook*. 1979–80 ed. Albany, N.Y.: Division of the Budget.

Petersilia, Joan. 1980. "Criminal Career Research: A Review of Recent Evidence." *Crime and Justice: An Annual Review of Research*, edited by Norval Morris and Michael Tonry. Vol. 2. Chicago: University of Chicago Press.

Peterson, Mark, and Harriet Braiker. 1980. *Doing Crime*. R-2200-DOJ. Santa Monica, Calif.: Rand.

Ponti, Gianluigi. 1962. "Studio statistico sulla criminalita senile." *Quaderni di criminologia clinica* 4:145–78.

Quetelet, Lambert A. J. 1969. *A Treatise on Man*. Gainesville, Fla.: Scholars' Facsimiles & Reprints.

Rangol, Alfred-Johannes. 1962. "Die Straffälligkeit nach Hauptdeliktsgruppen und Altersklassen 1884 bis 1958." *Monatschrift für Kriminologie und Strafrechtsreform* 15:157–75.

Reckless, Walter. 1973. *The Crime Problem*. Englewood Cliffs, N.J.: Prentice-Hall.

Rowe, Alan R., and Charles R. Tittle. 1977. "Life-Cycle Changes and Criminal Propensity." *Sociological Quarterly* 18:223–36.

Shannon, Lyle. 1978. "Predicting Adult Criminal Careers from Juvenile Careers." Paper presented at the 30th meeting of the American Society of Criminology, Dallas.

Siegel, Larry J., and Joseph J. Senna. 1981. *Juvenile Delinquency*. St. Paul, Minn.: West.

Sutherland, Edwin H., and Donald R. Cressey. 1970. *Criminology*, 8th ed. Philadelphia: Lippincott.

Swedish National Central Bureau of Statistics. 1980. *Yearbook of Legal Statistics, 1980*. Stockholm: National Central Bureau of Statistics.

Tittle, Charles R. 1980. *Sanctions and Social Deviance*. New York: Praeger.

Trasler, Gordon. 1980. "Aspects of Causality, Culture, and Crime." Paper presented at the 4th International Seminar at the International Centre of Sociological, Penal and Penitentiary Research and Studies, Messina, Sicily.

U.S. Department of Commerce, Bureau of the Census. 1893. *1890 Census of the Population: Special Schedule, Crime, Pauperism and Benevolence*. Washington, D.C.: Government Printing Office.

U.S. Department of Justice. 1979. *Uniform Crime Reports for the United States*. Washington, D.C.: Government Printing Office.

————. 1981–82. "Preliminary Analysis Yields New Data on Offender Activity." *Justice Assistance News* 2:10.

Wadsworth, Michael. 1979. *Roots of Delinquency*. New York: Harper & Row.

West, Donald J., and David Farrington. 1973. *Who Becomes Delinquent?* London: Heinemann.

————. 1977. *The Delinquent Way of Life*. London: Heinemann.

Wolfgang, Marvin. 1961. "Quantitative Analysis of Adjustment to the Prison Community." *Journal of Criminal Law, Criminology, and Police Science* 51:608–18.

Wolfgang, Marvin, Robert Figlio, and Thorsten Sellin. 1972. *Delinquency in a Birth Cohort*. Chicago: University of Chicago Press.

Wootton, Barbara. 1959. *Social Science and Social Policy*. New York: Macmillan.

Zimring, Franklin. 1981. "Kids, Groups and Crime: Some Implications of a Well-known Secret." *Journal of Criminal Law and Criminology* 72:867–85.

Zink, Theodore. 1958. "Are Prison Troublemakers Different?" *Journal of Criminal Law, Criminology, and Police Science* 48:433–34.

A Structural Decomposition of Black-White Earnings Differentials[1]

Robert L. Kaufman
University of Texas at Austin and *University of Utah*

This research demonstrates the impact of labor market structure and segmentation on differentials in black and white earnings. This paper argues that there are two very different factors which create black-white earnings differences: (1) differences between blacks and whites within divisions of the labor market, and (2) differences between labor market divisions in earnings combined with the differential distribution of blacks and whites across labor market divisions. Use of a decomposition based on a regression standardization approach discloses that the second factor accounts for a minimum of 14% of the black-white earnings gap. This implies that eliminating all black-white differences within labor market divisions would still leave a significant earnings gap between blacks and whites due to the structure of the labor market.

One of the predominant themes of the segmented labor market literature is that the structure of labor markets mediates the relationship between individual characteristics and rewards. Many researchers argue that the major impact of individual characteristics is on the sorting of workers into positions and that the major direct determinant of rewards is the labor market in which positions are located (see, e.g., Bluestone 1970; Stolzenberg 1975; Thurow 1975; Bibb and Form 1977; Wright 1978; Tolbert, Horan, and Beck 1980). A key implication is that discrimination in the sorting of individuals into positions should reproduce itself as discrimination in rewards.

Indeed, much of the motivation for the dual economy and dual labor market approaches was their use as an explanation of the persistent poverty and income differentials of racial-ethnic groups. These approaches argue that observed racial differences are partly or largely spurious since they are caused by the initial concentration of minorities in the lower-

wage periphery economy (secondary labor market) when they enter the labor force. Because employment in the periphery is inherently unstable, individuals in these groups will have unstable work histories. This instability is used by core sector (primary labor market) employers as evidence that minorities will not be good, steady employees. Consequently, they are blocked from moving into the higher-paying core jobs (Bluestone 1970; Doeringer and Piore 1971; Gordon 1972).

Writers outside of this school have come to similar conclusions concerning the impact of labor market segmentation. Stolzenberg (1975) analyzes racial differences in the effect of education on earnings within occupations and concludes that "the problem does not appear to be that black men are less able than white males *who work in the same occupation* to convert their schooling into earnings. Rather, it seems to be that the blacks are less successful than whites in converting their schooling into employment in better-paying occupations" (p. 314). This same logic has been employed in the race versus class debate. Wright (1978, p. 1369), for example, argues that ". . . if it is true that the returns to education vary substantially across class positions, and if it is true that black and white males are distributed quite differently across class positions, then much of the racial difference in returns to education could in fact be a consequence of the class distribution of races."

Despite the theoretical importance of this issue and the frequency with which it is used to explain between-group reward differences, there has been practically no research that compares the impact of rewards discrimination within labor markets with the impact of the differential distribution of social groups across labor market segments. Only Beck, Horan, and Tolbert (1980) have attempted to estimate the effect of the differential distribution of social groups across core and periphery economic sectors. They found, for both white and nonwhite females, that their actual distribution across sectors decreased their earnings compared with what they would earn if all groups had the same distribution across sectors. In contrast, for nonwhite males the differential distribution actually increased their earnings compared with what they would earn if there were no differential distribution. This rather contradictory finding is presented without comment or discussion by the authors.

One drawback to Beck et al.'s study is that the method that they use to estimate the impact of differential distribution does not permit a comparison of its effect with the effect of wage discrimination within labor market divisions. It may also misrepresent the influence of the differential distribution since it does not control for differences in levels of human capital and other individual characteristics across groups.

This is not surprising since a method for making such a comparison

had not yet been derived. One purpose of this paper is to present and apply such a method to estimate the impact of the differential distribution of groups across labor market divisions which can then be compared with the effect of rewards discrimination within labor market divisions. This method is an extension of the regression standardization approach used commonly by sociologists and economists to analyze group differences in occupational status and earnings (e.g., Duncan 1968; Hansen, Weisbrod, and Scanlon 1970; Winsborough and Dickinson 1971; Masters 1975; Featherman and Hauser 1976). The usual purpose of such a decomposition is to compare the relative impact of compositional effects (groups' differences in levels of human capital) and returns effects (groups' differences in returns to human capital). It is also used to compare the relative impacts of different sets of factors (e.g., family background vs. human capital). The decomposition that will be discussed later is unique in that it extends the comparison of compositional factors with returns factors in order to estimate the impact of the differential distribution of groups across labor market divisions.

DEFINING LABOR MARKET DIVISIONS

Most research on labor market segmentation has defined labor markets almost exclusively in terms of industrial location (Averitt 1968; Bluestone 1970; Bibb and Form 1977; Beck et al. 1978, 1980; Edwards 1979; Tolbert et al. 1980). But there has been growing criticism of such a definition because it ignores significant variation within industrial sectors in the composition and characteristics of work positions (Kalleberg, Wallace, and Althauser 1981; Schervish 1981, 1983; Wallace and Kalleberg 1981; Hodson and Kaufman 1982). That is, industrial sectors do not, as has often been claimed, contain homogeneous work positions; positions vary in terms of skill, routinization, authority, collective organization of workers, and other such factors (for evidence on this point see Hodson [1978, 1983]; Schervish [1981, 1983]; Wallace and Kalleberg [1981]).

This criticism is particularly relevant for research on minority groups since labor market discrimination against these groups operates not only in terms of differential allocation across industries but also in terms of differential allocation across work positions (Ashenfelter 1972; Shepherd 1970; Daymont and Kaufman 1979; Marshall 1974; Snyder and Hudis 1976; Kluegel 1978; Kaufman and Daymont 1981). Indeed, this may help explain why dual economy researchers have consistently found relatively little difference in the distribution of blacks and whites across the core and periphery industrial sectors (Hodson 1978; Beck et al. 1978, 1980; Tolbert et al. 1980; Zucker and Rosenstein 1981). It may be that the

587
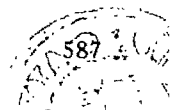
differential allocation of races is greater across work positions within sectors than it is across sectors as a whole.

For these reasons I define labor market divisions in terms of both industry and occupation by cross-classifying industrial sectors by occupational segments. I use a variation of the industrial sectors defined by Kaufman et al. (1981) and occupational segments as defined in my previous work (Kaufman 1981). Both sets of definitions are based on multidimensional cluster analyses of theoretically specified characteristics.

The industrial sectors are clusters of industries which are similar in terms of concentration, size, capital intensity, profitability, foreign involvement, growth, productivity, unionization, autonomy, and government ties. The original 16 sectors were reduced to 12 by merging the four single industry outlier clusters into the remaining 12.[2] In some ways these 12 sectors reproduce a core-versus-periphery distinction within broad industrial product types (e.g., core versus periphery utilities). But they cannot be grouped readily into global core and periphery sectors since many of the more detailed industrial sectors do not have consistent values on the defining characteristics (see Kaufman et al. [1981] for more details). In addition to these 12, the public sector is used as a separate industrial sector since there are well-documented differences between the public and private sectors in their treatment of minorities (Shepherd 1970; Smith and Welch 1977; Kaufman and Daymont 1981).

The occupational segments are clusters of occupations that are similar in terms of skill routinization, occupational task, unionization, and demand for labor. The eight segments can be divided globally into one set of high-skilled and less routinized occupations and another set of low-skilled and more routinized jobs. But this distinction is qualified within each set by the kinds of work tasks performed and the extent of unionization (see Kaufman 1981). Since both these factors are important determinants of minority employment opportunities (Snyder and Hudis 1976; Kluegel 1978; Franklin and Resnik 1976; Hill 1974; Marshall 1974; Kaufman and Daymont 1981), I use the more detailed eight segments. In addition, self-employed workers are treated as a separate occupational segment because these workers are subject to different economic pressures and conditions of work.

The cross-classification of the occupational segments with the industrial

---

[2] Tobacco was merged into the oligopoly sector since it is essentially an outlier higher than the oligopoly sector on most characteristics (see Kaufman, Hodson, and Fligstein 1981, table 4). Ordnance was collapsed into the core instead of into the oligopoly sector because it was closer to the core on most characteristics. The placement of brokers and real estate was difficult to specify since both have very irregular patterns of characteristics. They were placed in the periphery utilities sector since the cluster analysis results indicated that they were closest to this sector.

sectors defines the 116 labor market divisions used to apply the decomposition described below.[3]

A CONCEPTUAL DECOMPOSITION OF INTERGROUP
REWARD DIFFERENTIALS

Initially, intergroup reward differences can be thought of as having two components: an evaluation component and an allocation component (see Beck et al. 1980). The *evaluation* component represents the differential treatment of groups within labor market divisions in terms of overall rewards and the evaluation of their credentials. The *allocation* component represents the impact of the differential distribution of the groups across labor market divisions coupled with the differential rewards of employment in the various labor market divisions.

The evaluation component can be divided further into two parts. The *evaluation-composition* (EC) component represents differences in intergroup rewards which can be attributed to intergroup differences in levels of individual characteristics which affect rewards (e.g., intergroup differences in average education). The *evaluation-returns* (ER) component represents differences in rewards which can be attributed to different rates of return or payoffs to individual characteristics for the groups (e.g., different payoffs to education for the groups).

The meaning of these components can be illustrated by considering a two-sector (core-versus-periphery) decomposition of black-white earnings differences. The EC component represents the effect on earnings of differences between blacks and whites in the core sector in their levels of human capital, plus the effect of differences between blacks and whites in the periphery sector in their levels of human capital. The ER component represents the effect of differences between blacks and whites in the core sector in the payoff (return) to their human capital, plus the effect of differences between blacks and whites in the periphery sector in the payoff to their human capital.

Similarly, the allocation component can be divided into two parts. The *allocation-composition* (AC) component is differences in intergroup rewards that can be attributed to differences among labor market divisions of all worker characteristics combined with the different distribution of groups among labor market divisions. The *allocation-returns* (AR) component represents differences in intergroup rewards that can be attributed to differences among labor market divisions in the returns to worker characteristics combined with the different distribution of groups among

---

[3] The cross-classification of 13 industrial sectors by nine occupational segments gives 117 divisions. But with the use of census data it is impossible for a case to be coded as both self-employed and in the public sector, so there are no cases in this division.

589

labor market divisions. Let me reemphasize that these are differences among all workers in different divisions, not differences between groups.

In the two-sector example, the AC component would be the differences in black-white earnings attributable to differences between the core and periphery sectors in their mean levels of human capital, combined with the fact that blacks and whites have a different representation in the core and periphery sectors. The AR component would be the differences in black-white earnings attributable to differences between the core and periphery sectors in the payoffs to workers' human capital, combined with the differential distribution of blacks and whites in the core and periphery sectors.

There is a further decomposition that can be applied to the four components defined above. As shown in Appendix A, the probabilities of each group being employed in each labor market division are used to compute the components of the decomposition. But there are three different sets of probabilities that could be utilized for each group. One set is based on the observed (gross) distribution of each group among labor market divisions. The second set is based on a predicted net distribution of each group among labor market divisions, with standardization for differences between groups in levels of characteristics that affect their likelihood of employment in the labor market divisions. The third set of probabilities is equiprobable, based on assigning each group the same distribution among labor market divisions. Comparison of the results based on these three sets of probabilities permits an assessment of the impact of intergroup differences in employment likelihoods on rewards differences. The difference between the results using the gross and those using the net probabilities is the influence of intergroup differences in levels of characteristics that affect their employment opportunities. The difference between the results using the net and those using the equiprobable probabilities is the effect of intergroup discrimination in employment opportunities, net of intergroup differences in levels of characteristics that affect their employment opportunities.

Such a detailed decomposition allows the clear delineation of different sources of intergroup rewards differentials. It distinguishes between past discrimination (including non–labor market discrimination) and current labor market discrimination (i.e., EC vs. ER). It also permits an assessment of a key issue in the segmented labor market literature that is concerned with the impact of labor markets on discrimination. Comparing the ER with the AC and AR components provides an estimate of the relative sizes of the effects of discrimination in rewards within labor markets and discrimination in the distribution of groups among labor market divisions. This should make clear the compounding effect of discrimination in employment opportunities.

590

The mathematical model that corresponds to this conceptual decomposition is presented in Appendix A. In many ways it is a straightforward extension of the three-component decomposition which Winsborough and Dickinson (1971) present. As in their model there are composition, returns, and composition-returns interaction components. But my model differs from most decompositions in that the reference population (with which each group is compared) is the total population rather than one of the two groups. It also differs in its separation of an allocation from an evaluation component. It might prove enlightening to compare this detailed decomposition with the commonly used three-component decomposition which does not take into account the effects of labor market structure. Appendix B presents the mathematical model for this simpler decomposition.

OPERATIONALIZING THE DECOMPOSITION FOR BLACK-WHITE EARNINGS

The data for calculating the decomposition of black-white earnings differences were extracted from the 1970 U.S. Census Public Use Samples (PUS) (the state 5% and 15% samples and the neighborhood 5% and 15% samples). The sample of 1,502,386 whites and 130,290 blacks was restricted to the male experienced civilian labor force aged 20 or older in 1970. Also excluded were those whose 1970 industry or occupation code was "allocated" since it would not be possible to assign them to a labor market division. I chose to use these samples instead of other samples of the labor force with more extensive information on personal and labor force characteristics primarily because of sample size. Any other samples would not have sufficient blacks in many of the labor market divisions for reliable estimation. And the census samples contain sufficient information on personal and labor force characteristics for the analysis.[4]

Individual earnings were measured by totaling an individual's earnings (in hundreds of dollars) in 1969[5] from wages and salary, nonfarm business,

[4] The only drawback to the census samples is the known differential undercount of black and white males (Siegel 1974). The extent of the undercount for each race varies with age in virtually the same way; only the level of the undercount varies across race. Thus any comparison based on the absolute numbers of whites and blacks would be biased (e.g., the percentage of core workers who are white). But comparisons in relative terms within each race (e.g., the percentage of whites who are in the core with the percentage of blacks who are in the core), such as those focused on in this research, would not be biased necessarily by the different levels of the undercount of blacks and whites.

[5] The coding of cases into industrial sector and occupational segments is based on an individual's 1970 occupation and industry, but the measure of earnings is for the year 1969. Thus, the validity of using the 1970 U.S. Census PUS for this decomposition analysis is predicated on the assumption that the extent of between-division mobility over a single year's time is quite low. If there were a large amount of between-division mobility from 1969 to 1970, the mean earnings of the labor market divisions, and the decomposition results based on them, might be biased. But it is highly plausible that the extent of between-

and farm earnings. I used the sum of these three (rather than business earnings for the self-employed, farm earnings for farm owners, and wages and salary for all 'others) since the conditions of employment in different divisions of the labor market allow workers different opportunities for receiving earnings from these different sources. For example, many professionals can supplement their salaries by consultation work as part-time self-employed workers. Cases which reported no earnings from any of these sources were excluded from the analysis. To make the scores consistent, the figure for earnings from these three sources was truncated at $50,000 since the PUS tapes report each type of earnings truncated at $50,000.[6] The distribution of dollar earnings is highly skewed, so I will follow the common practice of using the natural logarithm of total earnings (in hundreds of dollars) as the dependent variable. Negative and zero earners were assigned a score of zero for the logged variable.

In order to apply the decomposition models presented in Appendixes A and B, a number of regression analyses of earnings are needed. The independent variables for the regression analyses include indicators of an individual's "human capital" and labor supply[7] as well as a set of indicators to control for regional and urban/rural variations in earnings. The inclusion of labor supply and regional controls is a debatable point. To some extent, racial differences in labor supply and in the effects of region are the result of labor market discrimination. My rationale for including them as controls is to be as conservative as possible in what is attributed to labor market discrimination. The equation which was estimated for the different groups in and across labor market divisions was

division mobility is low. Occupational shifts over a short period of time are most likely to occur between "related" occupations; i.e., between occupations sharing similar work tasks, training, etc. And industrial shifts over a short period of time are most likely to occur between related industries that structure work tasks similarly (Spilerman 1977). This would imply that there should be relatively little between-division mobility since the sectors and segments were defined by grouping together such related industries and occupations. In fact, I can estimate the amount of between-division mobility for roughly half of the sample from 1965 to 1970. Over this five-year period, 66.2% of the workers in the sample were in the same labor market division. This suggests that about 92.1% of the workers should be in the same division after one year. Thus there should be relatively little error involved in matching 1969 earnings with 1970 labor market division.

[6] The use of such a truncated dependent variable could potentially bias some of the regression results. However, it is unlikely to have much of an impact in this case, since the point at which truncation occurs is well over 2 standard deviations above the mean of earnings and since so few cases are involved (less than 0.5%).

[7] I did not include the census data's measure of number of hours worked per week for two reasons. First, this variable refers to the hours worked in 1970, not to those in 1969 when earnings is measured. More important, it refers to the hours worked during a particular reference week in 1970, not to typical hours worked per week. Weeks worked and indicators of work-related statuses five years previous should be a reasonable set of indicators of labor supply.

Log Earnings = $C_0 + C_1ED + C_2EDSQ + C_3EXP + C_4EXPSQ$
$+ C_5EDEXP + C_6WEEKS + C_7WORK65$
$+ C_8SCHOOL65 + C_9URBAN + C_{10}MID$-
ATLANTIC $+ C_{11}E.N.CENTRAL$ (1)
$+ C_{12}W.N.CENTRAL + C_{13}S.ATLANTIC$
$+ C_{14}E.S.CENTRAL + C_{15}W.S.CENTRAL$
$+ C_{16}MOUNTAIN + C_{17}PACIFIC.$

These independent variables were defined as follows:

ED   = number of years of schooling completed;
EDSQ   = $(ED)^2$;
EXP   = Age − ED − 6 as a proxy measurement of experience;
EXPSQ   = $(EXP)^2$;
EDEXP   = product of ED and EXP to allow for interaction effects;
WEEKS   = number of weeks worked in 1969;[1]
WORK65   = 1 if working at a job or business in 1965, 0 otherwise;
SCHOOL65   = 1 if attending school in 1965, 0 otherwise;
URBAN   = 1 if urban residence, 0 otherwise.
MID-ATLANTIC, E.N.CENTRAL, W.N.CENTRAL, S.ATLANTIC, E.S.CENTRAL, W.S.CENTRAL, MOUNTAIN, and PACIFIC are dummy variables for region of residence. An individual was given a code of 1 for the region in which he resided and a code of 0 for the other seven dummy variables. The New England region was used as the excluded category. (See U.S. Bureau of the Census [1972, p. 172] for the definition of these regions.)

As indicated in the Appendixes, equation (1) had to be estimated separately for the total sample, for blacks, and for whites within each of the labor market divisions as well as for these three groups pooled across labor market divisions. Of the 116 divisions, 16 had less than 30 blacks with nonmissing data, so these divisions were eliminated from the calculation of the decomposition.[9]

The decomposition also requires the computation of three sets of probabilities for blacks and whites of being employed in a particular labor market division to be used in conjunction with the within-division regression analyses. The observed (gross) probabilities for blacks and whites can be calculated directly from the sample data by

[1] The original coding of weeks worked in the census data set is an ordinal variable representing unequal ranges of weeks worked. This was converted to an interval scale by assigning the midpoint of the categories to each case. See U.S. Bureau of the Census (1972, p. 751) for the definition of the original variable.

[9] These 16 divisions were growers, professionals, and service in agriculture; professionals in core; growers in local monopoly; professionals in wholesale; self-employed in oligopoly; growers, professionals, and self-employed in core transport; growers in periphery transport; professionals in core utilities; technologists and professionals in periphery utilities; growers in small shops; professionals in periphery transport.

$$P_{\text{WI}} = \frac{\text{number of whites in division i}}{\text{total number of whites}} , \tag{2}$$

$$P_{\text{BI}} = \frac{\text{number of blacks in division i}}{\text{total number of blacks}} .$$

The equiprobable probabilities can also be calculated directly from the sample data, using the total number of persons in each labor market division to assign an equal probability to blacks and whites:

$$P_{\text{I}} = \frac{\text{number of whites and blacks in division i}}{\text{total number of whites and blacks}} . \tag{3}$$

The calculation of the net probabilities involved a more complex procedure which I will highlight briefly (details can be found in Kaufman [1981]). For each division of the labor market, a set of log-linear analyses was performed to estimate the net effect of race on the likelihood (odds) that a worker was employed in that labor market division. Controls were included for age, education, region, residential location, employment level (part-time vs. full-time), and labor force participation five years previous. Based on these analyses, an average net black-white employment likelihood was calculated for each labor market division. This procedure is analogous to performing a set of regression analyses to predict the probability that a person of each race is in a given labor market division, net of the effects of age, education, region, residential location, employment level, and labor force participation five years previous. The net black-white employment likelihood is analogous to the predicted odds, based on the regression equation for the division, that a black compared with a white with equivalent characteristics (i.e., equal mean age, education, etc.) is in the labor market division. Table 1 presents these average net employment likelihoods for each combination of industrial sector and occupational segment. Odds ratios or ratios of odds ratios can be constructed from this table to describe the relationship of race to employment in the sectors and segments (see Bishop, Fienberg, and Holland 1975; Daymont and Kaufman 1979).

Table 1 is quite similar to the table that is needed to calculate the net probabilities. What is needed is a table that has the same pattern of odds ratios for race as table 1 but that has cell entries scaled to be numbers of cases instead of odds. Such a table can be derived from table 1 by Demming-adjusting it to a specified set of margins: the one-way marginal distribution of race and the two-way marginal distribution of sector by segment for the sample data (see Bishop et al. [1975] for more discussion of Demming-adjustments). This procedure redistributes blacks and whites

TABLE 1

AVERAGE WITHIN-DIVISION BLACK-WHITE EMPLOYMENT LIKELIHOODS (Odds)

INDUSTRIAL SECTOR	OCCUPATIONAL SEGMENTS								
	Craft	Technologist	Grower	Coordinator	Professional	Service	Worker	Secondary	Self-Employed
Agriculture	1.000	.943	.579	.651	.874	1.506	1.201	1.583	.664
Periphery	.781	.700	.637	.593	.823	.922	1.201	1.153	.687
Core	.753	.572	.582	.521	.234	1.173	1.134	1.214	.539
Local monopoly	.661	.891	.546	.618	.880	1.281	1.256	1.844	.677
Wholesale	.671	.599	.635	.452	.386	1.090	1.053	1.147	.458
Oligopoly	1.047	.766	.671	.692	.408	1.528	1.476	1.334	1.024
Core transport	.627	.681	.740	.555	.251	1.230	1.173	1.147	.750
Periphery transport	.834	.778	.305	.524	.378	.856	.949	1.257	.901
Core utility	.624	.796	.684	.560	.479	.832	1.178	1.633	.428
Periphery utility	.951	.904	.536	.638	.449	1.257	1.563	1.445	.619
Small shop	.681	.887	.558	.558	.845	.704	1.098	1.318	.554
Education and nonprofit	.904	1.261	1.232	.988	1.092	1.295	1.459	2.146	1.047
Public	.796	.914	1.000	.916	1.388	1.583	1.510	1.543	...

among labor market divisions to reflect their net employment likelihoods (odds), while it preserves the same total number of blacks and whites across divisions and the same total number of cases in each division. Table 2 presents the Demming-adjusted version of table 1. From this table the net probabilities for whites and blacks can be calculated as

$$P_{\mathrm{W}i}{}^* = \frac{\text{predicted number of whites in division } i}{\text{total number of whites}} ,$$

$$P_{\mathrm{B}i}{}^* = \frac{\text{predicted number of blacks in division } i}{\text{total number of blacks}} .$$

(4)

Before I discuss the results of the various decompositions, I should note two further cautions concerning possible biases. First, Olneck (1976) has shown that using education and its square as predictors of earnings instead of using a functional form that takes account of nonlinearities due to credentialism may create some overestimation of the return component in decomposing earnings using regression. Since the analysis presented here uses education and its square, there may be some upward bias in the estimation of the returns components. Second, it is quite likely that there would be some correlation between the errors in the earnings regressions for each labor market division and the errors in the log-linear analyses for each labor market division which were used to measure the net probabilities. The rationale is that it is likely that those employers who discriminate in earnings will also discriminate in hiring and promotion and vice versa. But it is very difficult to say exactly how this problem of correlated errors would affect the analysis presented here, because this does not fit the typical situation in which correlated errors create bias (i.e., we are not regressing earnings on the black-white likelihood or vice versa).

DECOMPOSITION RESULTS

Table 3 presents the earnings regression equations and variable means that were used to compute the decomposition which does not take the structure of the labor market into account. Table 4 presents the results of this standard decomposition. Overall, there is a gross difference of .4812 between the white and black means of log earnings which is equivalent to a dollar differential of $2,770 in 1969 dollars.[10] Of this total difference, 44% can be attributed to compositional differences between

[10] This dollar conversion was calculated by taking the dollar equivalent of white mean log earnings minus the dollar equivalent of black mean log earnings. It should be remembered that log earnings was defined as the natural logarithm of earnings in hundreds of dollars.

TABLE 2

DEMMING-ADJUSTED TABLE OF NUMBER OF BLACKS AND WHITES IN LABOR MARKET DIVISIONS

INDUSTRIAL SECTOR	Craft	Technologist	Grower	Coordinator	Professional	Service	Worker	Secondary	Self-Employed
Agriculture	68	123	5	124	23	3	215	3,021	3,247
	689	1,323	84	1,928	270	19	1,810	19,308	49,460
Periphery	2,254	1,864	281	2,382	216	1,101	7,303	752	2,257
	29,214	27,110	4,457	40,642	2,661	12,071	61,523	6,595	33,255
Core	3,542	1,099	217	2,620	58	169	11,935	531	339
	47,543	19,455	3,774	50,874	2,515	1,455	106,557	4,424	6,366
Local monopoly	3,739	380	14	818	258	459	2,823	370	1,849
	57,206	4,312	268	13,387	2,974	3,625	22,679	2,030	27,594
Wholesale	1,421	846	114	2,132	185	60	4,222	338	1,005
	21,447	14,295	1,810	47,796	4,857	553	40,560	2,984	22,159
Oligopoly	1,065	530	139	735	23	29	3,044	87	35
	10,301	7,005	2,018	10,737	574	194	20,854	660	348
Core transport	1,256	319	44	509	24	127	1,374	693	48
	20,279	4,739	596	9,283	948	1,046	11,856	6,121	653
Periphery transport	60	127	2	226	5	66	2,236	116	387
	729	1,650	51	4,362	141	775	23,868	932	4,354
Core utility	472	233	152	2,001	52	164	834	69	149
	7,659	2,969	2,251	36,185	1,098	1,999	7,175	427	3,519
Periphery utility	126	33	146	524	11	87	597	181	351
	1,341	364	2,761	8,310	245	703	3,867	1,265	5,724
Small shop	789	1,837	16	1,829	128	2,715	3,566	368	2,563
	11,719	20,924	289	33,196	1,534	38,993	32,822	2,820	46,791
Education and nonprofit	155	145	117	350	2,054	184	874	565	244
	1,737	1,168	958	3,582	19,033	1,435	6,066	2,662	2,360
Public	2,042	1,637	668	4,614	5,730	1,683	7,273	1,883	...
	25,969	18,138	6,760	50,980	41,767	10,752	48,725	12,358	...

NOTE.—Cell entries are number of blacks on first line and number of whites on second line.

TABLE 3

EARNINGS DECOMPOSITION REGRESSION EQUATIONS AND VARIABLE MEANS FOR TOTAL, WHITE, AND BLACK SAMPLES POOLED ACROSS LABOR MARKET DIVISIONS

VARIABLE*	TOTAL SAMPLE			WHITE SAMPLE			BLACK SAMPLE		
	b	SE	Mean	b	SE	Mean	b	SE	Mean
Constant	1.060	...	...	1.162	...	...	1.125	...	...
MID-ATLANTIC	.014	.0025	.193	.026	.0025	.194	−.023	.0152	.179
E.N.CENTRAL	.046	.0025	.209	.054	.0025	.212	.056	.0152	.183
W.N.CENTRAL	−.124	.0029	.083	−.120	.0029	.088	−.139	.0181	.029
S.ATLANTIC	−.127	.0026	.144	−.075	.0027	.131	−.245	.0151	.288
E.S.CENTRAL	−.206	.0031	.060	−.159	.0032	.056	−.364	.0158	.104
W.S.CENTRAL	−.173	.0028	.092	−.140	.0029	.089	−.325	.0155	.126
MOUNTAIN	−.133	.0036	.035	−.133	.0036	.037	−.161	.0274	.006
PACIFIC	.004	.0026	.128	.010	.0027	.133	−.021	.0161	.070
URBAN	.151	.0012	.736	.163	.0013	.729	.246	.0053	.824
WEEKS	.037	.0001	47.938	.037	.0001	48.049	.034	.0002	46.662
WORK65	.204	.0018	.872	.207	.0019	.874	.138	.0055	.841
SCHOOL65	−.054	.0028	.053	−.056	.0028	.055	.012	.0124	.027
ED	.039	.0011	11.529	.026	.0012	11.700	.037	.0036	9.559
EDSQ	.002	.0000	144.781	.002	.0000	148.202	.002	.0001	105.320
EXP	.047	.0003	24.679	.047	.0003	24.598	.040	.0010	25.608
EXPSQ	−.001	.0000	822.364	−.001	.0000	817.153	−.001	.0000	882.469
EDEXP	−.001	.0000	261.009	−.001	.0000	265.308	−.001	.0001	211.420

* See text for explanation of variables.

blacks and whites and 59% to return differentials between blacks and whites.[11] This model implies that a substantial portion of the black-white earnings difference is due to differences between blacks and whites in their levels of human capital, labor supply, and residential locations. But it also suggests that the lion's share of the total difference lies in earnings discrimination in terms of differential payoffs to characteristics for blacks and whites. Let me reiterate that the latter figure is in many ways a conservative estimate of the effect of discrimination against blacks in the labor market since market discrimination causes some part of the black-white differences in levels of labor supply.

A different picture emerges from a consideration of the more detailed decomposition results[12] presented in table 5. Part A of table 5 presents the decomposition based on the observed (gross) employment probabilities,[13] part B presents the results based on the net employment likelihoods, and part C presents the results based on the equiprobable probabilities (i.e., the assumption of no differential distribution of blacks and whites among labor market divisions). Table 6 presents contrasts among these sets of results. Part A of table 6 shows the difference between the gross and net probability results which represents the effect on the earnings gap of black-white differences in levels of characteristics which affect their employment opportunities. Part B of table 6 shows the difference between the net and the equiprobable probability results which represents

TABLE 4

DECOMPOSITION OF BLACK-WHITE EARNINGS DIFFERENCE
IGNORING LABOR MARKET STRUCTURE

	Log Earnings	% of Total Difference
Effect of black-white composition	.2092	43.5
Effect of black-white returns	.2841	59.0
Interaction of composition and returns	−.0121	−2.5
Total difference	.4812	100.0

NOTE.—These figures were calculated using equation (B3) from Appendix B and the information presented in table 3.

[11] The tables presenting the regression results and variables means for the 116 divisions on which table 5 is based are not included because of their bulk (70 pages) but will be made available by the author for the cost of reproduction and mailing.

[12] Although there is a term in the decomposition for the interaction of composition and returns, the term is so small relative to the main effects of composition and returns that I will simplify the discussion by ignoring it.

[13] The small divergence in the total difference in log earnings between tables 4 and 5 is due to the exclusion in the full decomposition of several divisions whose cases were included in the computation of the decomposition presented in table 4.

the effect on the earnings gap of discrimination in employment opportunities.

The decomposition in part A of table 5 suggests that black-white differences in levels of human capital, labor supply, region, and residence (the EC component) account for only 26% of the total black-white earnings

TABLE 5

FULL DECOMPOSITION OF BLACK-WHITE EARNINGS DIFFERENCE

A. BASED ON OBSERVED (Gross) PROBABILITIES

	Log Earnings	% of Total Difference
Total allocation component...................	.1750	36.6
Allocation-composition....................	.0734	15.3
Allocation-returns........................	.1016	21.2
Total evaluation component	.3038	63.4
Evaluation-composition...................	.1227	25.6
Evaluation-returns.......................	.1926	40.2
Interaction of composition and returns......	−.0115	−2.4
Total difference........................	.4788	100.0

B. BASED ON NET PROBABILITIES

	Log Earnings	% of Total	% of Panel Total
Total allocation component...................	.0672	14.0	16.6
Allocation-composition....................	.0250	5.2	6.2
Allocation-returns........................	.0422	8.8	10.4
Total evaluation component	.3387	70.7	83.4
Evaluation-composition...................	.1352	28.2	33.3
Evaluation-returns.......................	.2146	44.8	52.9
Interaction of composition and returns......	−.0111	−2.3	−2.7
Total difference........................	.4058	84.8	100.0

C. BASED ON EQUIPROBABLE PROBABILITIES

	Log Earnings	% of Total	% of Panel Total
Total allocation component...................	.0000	.0	.0
Total evaluation component	.3758	78.5	100.0
Evaluation-composition...................	.1507	28.9	40.1
Evaluation-returns.......................	.2367	49.6	63.2
Interaction of composition and returns......	−.0125	−2.6	−3.3
Total difference........................	.3758	78.5	100.0

NOTE.—The numbers in parts A, B, and C are based on using the observed, net, and equiprobable probabilities, respectively, in eqq. (A9)–(A11), (A14), and (A15) of Appendix A.

differential. This contrasts sharply with the 44% figure implied by the standard decomposition. Similarly the structural decomposition model implies that 40% of the black-white earnings gap is due to differences in returns to characteristics for blacks and whites within divisions, compared with the 59% implied by the standard decomposition.

Perhaps the most important result from table 5, part A, is the size of the allocation component which accounts for the differences noted above between the standard and the structural decompositions. The differential composition of workers within divisions explains 15% of the total black-white gap, while the differential returns to characteristics within divisions accounts for the remaining 21%. It should be reemphasized that these are between-division differentials, not black-white differences, which affect the total black-white earnings difference through the unequal distribution of blacks and whites among divisions. These results suggest that, overall, 37% of the black-white earnings gap can be attributed to the effect of different wage determination processes among divisions and the

TABLE 6

DECOMPOSING THE EFFECTS OF COMPOSITION AND DISCRIMINATION
ON EMPLOYMENT OPPORTUNITIES

	Log Earnings	% of Total	% of Panel Total
A. Due to black-white composition effects on employment opportunities			
Total allocation component..................	.1078	22.5	147.7
Allocation-composition....................	.0484	10.1	66.3
Allocation-returns.......................	.0594	12.4	81.4
Total evaluation component	−.0349	−7.3	−48.8
Evaluation-composition...................	−.0125	−2.6	−17.1
Evaluation-returns.......................	−.0220	−4.6	−30.1
Interaction of composition and returns......	−.0004	−.0	−.5
Total difference.......................	.0730	15.2	100.0
B. Due to discrimination in employment opportunities			
Total allocation component..................	.0672	14.0	224.0
Allocation-composition....................	.0250	5.2	83.3
Allocation-returns.......................	.0422	8.8	140.7
Total evaluation component	−.0371	−7.7	−123.7
Evaluation-composition...................	−.0155	−3.2	−51.7
Evaluation-returns.......................	−.0230	−4.8	−76.7
Interaction of composition and returns......	.0014	.3	4.7
Total difference.......................	.0300	6.3	100.0

NOTE.—Part A is based on table 5, part A minus part B, and part B is based on table 5, part A minus part C.

differential distribution of blacks and whites among divisions. However, this can be decomposed further into two components. The first reflects the effect on the black-white earnings gap of discrimination in black-white employment opportunities. The second component represents the influence on this earnings gap of the two groups' differential composition as it affects their employment opportunities. Table 6 suggests that 38% of the total allocation component can be attributed to discrimination in employment opportunities (.0672 of .1750), while the remaining 62% can be attributed to the effects of differential black-white composition on employment opportunities (.1078 of .1750).

The total evaluation component accounts for 63% of the total black-white earnings gap (table 5, part A). Decomposing this component into the impact of discrimination in employment opportunities and composition effects on employment opportunities produces a somewhat curious result. The component representing the influence of composition differentials on employment opportunities is negative (table 6, part A). This suggests that, if it were possible to eliminate the effects of black-white composition on employment opportunities without making any other changes in labor market processes, there would be even greater differences in black-white earnings due to within-division black-white differences (the evaluation component), although the total black-white earnings gap would decrease (compare the total difference in part A of table 5 with part A of table 6 and the total evaluation component in part A of table 5 with part A of table 6). The implication of this result is that if unequal black-white employment opportunities were evened out by eliminating compositional differences, more blacks would move into labor market divisions having greater within-division black-white earnings differences. There is a similar result with the use of the assumption that all differences in black-white employment opportunities are eliminated (compare parts A and C of table 5). Again, the absolute size of the evaluation component increases, although the total black-white earnings gap decreases.

These results are consistent with Beck et al.'s (1980) finding of a decline in nonwhite male earnings with the assumption of no allocation differentials. Moreover, their results also show that the male white-nonwhite earnings gap decreases with the assumption of no allocation differences (compare the actual and revised figures in their table 2), even though nonwhite males' earnings decline in absolute terms. Although these results are a logical implication of segmented labor market theories, the implication has not been arrived at before. Dual economy researchers, for example, have argued that not only are blacks kept out of the core sector but also there is greater discrimination against blacks in the core (Doeringer and Piore 1971; Gordon 1972; Beck et al. 1980). Putting these two

arguments together implies that eliminating unequal employmen: opportunities should move more blacks into the core sector where they would then be facing greater wage discrimination. This is precisely the kind of result indicated by the structural decomposition presented above.

These decomposition results can also be used to speculate about the impact and effectiveness of various social policies aimed at reducing racial earnings differentials. First of all, consider what would be implied by the elimination of compositional differences between blacks and whites. Under the structural decomposition model there are several different components which represent the impact of differential black-white composition. Elimination of black-white compositional effects on employment opportunities (table 6, part A) would create an overall reduction in black-white earnings differentials of 15%. (But this represents a 23% reduction due to between-division allocation effects and a 7% increase due to within-division evaluation effects.) Above and beyond this, there is an additional 28% reduction in black-white earnings differences due to the elimination of within-division black-white composition differences (the EC component in table 5, part B). This suggests that the total effect of eliminating differential black-white composition would be to reduce the black-white earnings gap by 43%. While this is a sizable difference, the fact that it is still less than half of the total black-white earnings gap points to the inadequacy of social planning and policy focusing solely on "upgrading" minority job skills and resources.

It is also possible to evaluate the effect of eliminating all differences in black-white employment opportunities by comparing parts A and C of table 5. These results suggest that equalizing black-white employment opportunities would lead to an overall reduction of 22% of the total black-white earnings gap. But this reflects a 37% decrease due to the elimination of allocation effects, while the impact of the evaluation component is to increase black-white earnings differences by 15%. There is a similar result if only discrimination in employment opportunities is eliminated. Table 6, part B, indicates a total reduction of the black-white earnings gap of 6%, reflecting a 14% decrease due to allocation effects but an 8% increase due to evaluation effects. Thus, the evening out of differences in black-white employment opportunities would reduce the earnings gap between blacks and whites, but it would also move blacks into parts of the labor market where, apparently, there is greater discrimination against blacks in wages.

The standard and the structural decomposition models have somewhat different implications for social planning. This can be seen most easily by comparing the prediction of the two models in the situation in which all black-white compositional differences have been eliminated and black-

white. earnings returns differences have also been eliminated. In this situation the standard decomposition model would predict the complete elimination of the black-white earnings gap. But the structural decomposition model would predict only an 86% reduction; the allocation component representing discrimination in employment opportunities would remain (table 6, part B). Thus, the most conservative estimate of the impact of discrimination in employment opportunities suggests that it accounts for 14% of the total black-white earnings gap.

CONCLUSIONS

Previous theorizing and research has long speculated on the importance of the unequal distribution of races among labor market divisions as an explanation of racial wage inequalities. The results presented here have confirmed this speculation. Overall, the differential distribution of blacks and whites among labor market divisions explains over one-third of the total black-white earnings gap. Much of this factor represents the direct impact of labor market discrimination in employment opportunities, but most of it is the result of other kinds of discrimination (e.g., in education or within regions). While it is clear that this factor has a substantial effect, it is equally clear that most of the black-white earnings gap is caused by differences within labor market divisions in wage determination processes for blacks and whites. In this instance at least, race appears to be as important as the labor market structure in determining the outcome for blacks.

These results also emphasize the importance of social policy planning which focuses on inequalities in black-white employment opportunities, not just on differences in levels of, or returns to, "human capital" factors. Policy and planning which do not take into account the structuring of the labor market cannot succeed in eliminating the racial wage gap. The elimination of racial wage differences has no single solution. It requires efforts to eliminate differences in access to skills and training, efforts to eliminate differential payoffs for such factors, and efforts to eliminate the unequal distribution of blacks and whites among the structural divisions of the labor market.

The results of the structural decomposition have also highlighted an interesting implication of segmented labor market theories. These approaches have argued that minorities are distributed differently among labor markets and that there are comparable differences within labor markets in the extent of wage discrimination against minorities. But the logical implication of these arguments has never been drawn out: the elimination of unequal employment opportunities should move minorities

into labor markets which practice greater wage discrimination. The decomposition results presented here suggest that this would result in a net gain in earnings for black males, even though they would be subject to greater wage discrimination within labor market divisions.

Given the contradictory findings of Beck et al. (1980) concerning black versus female earnings using the assumption of no differential distribution, it may be that eliminating the differential distribution of females among labor market divisions would not involve the same trade-offs that appear for black males. It is to be hoped that the structural decomposition model presented here will prove useful for such future investigations of the minority group and labor market structure relationship. This model should also prove readily generalizable to other similar research concerns.

APPENDIX A

Defining the Components of the Structural Decomposition Model

To develop the mathematical model, I begin with the total intergroup rewards difference measured by the group 1 mean of rewards ($\overline{Y}_1$) minus the group 2 mean of rewards ($\overline{Y}_2$).

$$
\begin{aligned}
D &= \overline{Y}_1 - \overline{Y}_2 \\
&= (\overline{Y}_1 - \overline{Y}) - (\overline{Y}_2 - \overline{Y}) ,
\end{aligned}
\tag{A1}
$$

where $\overline{Y}$ is the total (groups 1 and 2) mean of rewards. But $\overline{Y}_1$, $\overline{Y}_2$, and $\overline{Y}$ can be written in terms of the corresponding means in each division of the labor market and the probabilities for each group of being in each division of the labor market. That is,

$$
\begin{aligned}
\overline{Y}_1 &= \sum_i \overline{Y}_{1i} P_{1i} , \\
\overline{Y}_2 &= \sum_i \overline{Y}_{2i} P_{2i} , \\
\overline{Y} &= \sum_i \overline{Y}_i P_i ;
\end{aligned}
\tag{A2}
$$

where $\overline{Y}_{1i}$, $\overline{Y}_{2i}$, and $\overline{Y}_i$ are the respective mean rewards of group 1, group 2, and the total in labor market division i; and P_{1i}, P_{2i}, and P_i are the respective probabilities for group 1, group 2, and the total of being in labor market division i.

Substituting from (A2) into (A1) gives

$$D = \left[\sum_i (P_{1i}\overline{Y}_{1i} - P_i\overline{Y}_i)\right] - \left[\sum_i (P_{2i}\overline{Y}_{2i} - P_i\overline{Y}_i)\right]$$

$$= \left\{\sum_i [P_{1i}(\overline{Y}_{1i} - \overline{Y}_i) + (P_{1i} - P_i)\overline{Y}_i]\right\}$$

$$- \left\{\sum_i [P_{2i}(\overline{Y}_{2i} - \overline{Y}_i) + (P_{2i} - P_i)\overline{Y}_i]\right\} \tag{A3}$$

$$= \sum_i P_{1i}(\overline{Y}_{1i} - \overline{Y}_i) - \sum_i P_{2i}(\overline{Y}_{2i} - \overline{Y}_i) + \sum_i (P_{1i} - P_{2i})\overline{Y}_i .$$

The first two terms in the last line of (A3) correspond to the evaluation component (D_E) since they represent the difference in intergroup rewards which is caused by different rewards for the groups within a labor market division, contrasted with the total division mean of rewards. The last term represents the allocation component (D_A) since it reflects the difference in intergroup rewards which is due to the differential distribution of the groups among divisions coupled with the different mean rewards of the divisions. Thus,

$$D_E = \sum_i P_{1i}(\overline{Y}_{1i} - \overline{Y}_i) - \sum_i P_{2i}(\overline{Y}_{2i} - \overline{Y}_i) , \tag{A4}$$

$$D_A = \sum_i (P_{1i} - P_{2i})\overline{Y}_i . \tag{A5}$$

The next step is to decompose both D_E and D_A into components which represent compositional differences and returns differentials. This requires a set of regressions of rewards on a set of characteristics (X) within each labor market division for group 1, group 2, and the total as well as a regression for the total group pooled across all of the labor market divisions. The various rewards means can then be written in terms of the regression coefficients and the independent variable means as

$$\overline{Y}_{1i} = \sum_k C_{1ik}\overline{X}_{1ik} ,$$

$$\overline{Y}_{2i} = \sum_k C_{2ik}\overline{X}_{2ik} ,$$

$$\overline{Y}_i = \sum_k C_{ik}\overline{X}_{ik} , \tag{A6}$$

$$\overline{Y} = \sum_k C_k\overline{X}_k ;$$

where the C_{1ik}, C_{2ik}, C_{ik}, and C_k are the unstandardized regression coefficients from the three sets of within-division analyses and from the total

pooled analysis, respectively; and the $\overline{X}_{1ik}$, $\overline{X}_{2ik}$, $\overline{X}_{ik}$, and $\overline{X}_k$ are the means for the independent variables for group 1, group 2, and total within-division i and for the total pooled sample.

Substituting from (A6) into (A4) gives

$$
\begin{aligned}
D_E &= \sum_i P_{1i}\left(\sum_k C_{1ik}\overline{X}_{1ik} - \sum_k C_{ik}\overline{X}_{ik}\right) \\
&\quad - \sum_i P_{2i}\left(\sum_k C_{2ik}\overline{X}_{2ik} - \sum_k C_{ik}\overline{X}_{ik}\right) \\
&= \sum_i P_{1i}\left[\sum_k C_{1ik}(\overline{X}_{1ik} - \overline{X}_{ik}) + \sum_k (C_{1ik} - C_{ik})\overline{X}_{ik}\right] \\
&\quad - \sum_i P_{2i}\left[\sum_k C_{2ik}(\overline{X}_{2ik} - \overline{X}_{ik}) + \sum_k (C_{2ik} - C_{ik})\overline{X}_{ik}\right] .
\end{aligned}
\tag{A7}
$$

Rearranging terms in (A7) gives

$$
\begin{aligned}
D_E &= \sum_i \left[P_{1i}\sum_k C_{ik}(\overline{X}_{1ik} - \overline{X}_{ik}) - P_{2i}\sum_k C_{ik}(\overline{X}_{2ik} - \overline{X}_{ik})\right] \\
&\quad + \sum_i \left[P_{1i}\sum_k (C_{1ik} - C_{ik})\overline{X}_{ik} - P_{2i}\sum_k (C_{2ik} - C_{ik})\overline{X}_{ik}\right] \\
&\quad + \sum_i \left[P_{1i}\sum_k (C_{1ik} - C_{ik})(\overline{X}_{1ik} - \overline{X}_{ik}) \right. \\
&\quad \left. - P_{2i}\sum_k (C_{2ik} - C_{ik})(\overline{X}_{2ik} - \overline{X}_{ik})\right] .
\end{aligned}
\tag{A8}
$$

The first term in (A8) corresponds to the EC component (D_{EC}) that was discussed in the text, since it reflects the differences in the means of characteristics between groups within a labor market division, using a contrast of each group to the total within-division means of the characteristics. The second term in (A8) corresponds to the ER component (D_{ER}) discussed above, since it reflects the impact of unequal returns to characteristics for the two groups, using a contrast of each group to the total within-division returns to characteristics. The third term in (A8) is an interaction component (D_{ECR}) which represents heuristically the effect of simultaneously changing intergroup composition and returns (see Wins-borough and Dickinson [1971] for more discussion). Writing out each of these three components separately gives

$$
D_{EC} = \sum_i \left[P_{1i}\sum_k C_{ik}(\overline{X}_{1ik} - \overline{X}_{ik}) - P_{2i}\sum_k C_{ik}(\overline{X}_{2ik} - \overline{X}_{ik})\right] ,
\tag{A9}
$$

$$D_{ER} = \sum_i \left[P_{1i} \sum_k (C_{1ik} - C_{ik})\overline{X}_{ik} - P_{2i}\sum_k (C_{2ik} - C_{ik})\overline{X}_{ik} \right], \quad \text{(A10)}$$

$$D_{ECR} = \sum_i \left[P_{1i} \sum_k (C_{1ik} - C_{ik})(\overline{X}_{1ik} - \overline{X}_{ik}) \right.$$
$$\left. - P_{2i}\sum_k (C_{2ik} - C_{ik})(\overline{X}_{2ik} - \overline{X}_{ik}) \right]. \quad \text{(A11)}$$

The allocation component (D_A) can also be decomposed further by an additional expansion in terms of differences in means.

$$D_A = \sum_i (P_{1i} - P_{2i})(\overline{Y}_i - \overline{Y}) + \sum_i (P_{1i} - P_{2i})\overline{Y}$$
$$= \sum_i (P_{1i} - P_{2i})(\overline{Y}_i - \overline{Y}) \quad \text{(A12)}$$
$$\text{since } \sum_i (P_{1i} - P_{2i})\overline{Y} = \overline{Y}\left(\sum_i P_{1i} - \sum_i P_{2i}\right) = \overline{Y}(1 - 1) = 0.$$

Substituting from (A6) into (A12) gives

$$D_A = \sum_i (P_{1i} - P_{2i})\left(\sum_k C_{ik}\overline{X}_{ik} - \sum_k C_k\overline{X}_k\right)$$
$$= \sum_i (P_{1i} - P_{2i})\left[\sum_k C_{ik}(\overline{X}_{ik} - \overline{X}_k) + \sum_k (C_{ik} - C_k)\overline{X}_k\right] \quad \text{(A13)}$$
$$= \sum_i (P_{1i} - P_{2i})\sum_k C_{ik}(\overline{X}_{ik} - \overline{X}_k)$$
$$+ \sum_i (P_{1i} - P_{2i})\sum_k (C_{ik} - C_k)\overline{X}_k.$$

The first term in (A12) corresponds to the AC component discussed in the text (D_{AC}), while the second term corresponds to the AR component (D_{AR}). That is,

$$D_{AC} = \sum_i (P_{1i} - P_{2i})\sum_k C_{ik}(\overline{X}_{ik} - \overline{X}_k), \quad \text{(A14)}$$

$$D_{AR} = \sum_i (P_{1i} - P_{2i})\sum_k (C_{ik} - C_k)\overline{X}_k. \quad \text{(A15)}$$

APPENDIX B

Defining the Components of the Standard Decomposition

608

This decomposition also starts with the total between-group difference in mean rewards:

$$D = (\overline{Y}_1 - \overline{Y}) - (\overline{Y}_2 - \overline{Y}) . \qquad \text{(B1)}$$

The next step requires results from a set of regressions of rewards on a set of characteristics (X) for group 1, group 2, and the total sample. Then the rewards means in (B1) can be written in terms of the unstandardized regression coefficients and the independent variable means:

$$\overline{Y}_1 = \sum_k C_{1k}\overline{X}_{1k} ,$$

$$\overline{Y}_2 = \sum_k C_{2k}\overline{X}_{2k} , \qquad \text{(B2)}$$

$$\overline{Y} = \sum_k C_k \overline{X}_k ;$$

where the C_{1k}, C_{2k}, and C_k are the unstandardized regression coefficients from the group 1, group 2, and total regression, respectively, and the $\overline{X}_{1k}$, $\overline{X}_{2k}$, and $\overline{X}_k$ are the means of the independent variables for group 1, group 2, and total samples, respectively.

Substituting from (B2) into (B1) gives

$$D = \sum_k (C_{1k}\overline{X}_{1k} - C_k\overline{X}_k) - \sum_k (C_{2k}\overline{X}_{2k} - C_k\overline{X}_k)$$

$$= \sum_k [C_{1k}(\overline{X}_{1k} - \overline{X}_k) + (C_{1k} - C_k)\overline{X}_k]$$

$$- \sum_k [C_{2k}(\overline{X}_{2k} - \overline{X}_k) + (C_{2k} - C_k)\overline{X}_k]$$

$$= \left[\sum_k C_k(\overline{X}_{1k} - \overline{X}_k) - \sum_k C_k(\overline{X}_{2k} - \overline{X}_k) \right]$$

$$+ \sum_k (C_{1k} - C_{2k})\overline{X}_k \qquad \text{(B3)}$$

$$+ \left[\sum_k (C_{1k} - C_k)(\overline{X}_{1k} - \overline{X}_k) - \sum_k (C_{2k} - C_k)(\overline{X}_{2k} - \overline{X}_k) \right]$$

$$= \sum_k C_k(\overline{X}_{1k} - \overline{X}_{2k})$$

$$+ \sum_k (C_{1k} - C_{2k})\overline{X}_k$$

$$+ \left[\sum_k (C_{1k} - C_k)(\overline{X}_{1k} - \overline{X}_k) - \sum_k (C_{2k} - C_k)(\overline{X}_{2k} - \overline{X}_k) \right] .$$

609

The first line in (B3) represents the effects of differential composition (D_C), the second line represents the impact of differential returns (D_R), and the last line represents the interaction effect of simultaneously changing composition and returns between groups. This model exactly parallels Winsborough and Dickinson's (1971) three-component decomposition, except that, as the reference population, it uses the total group instead of one of the two groups.

REFERENCES

Ashenfelter, O. 1972. "Racial Discrimination and Trade Unionism." *Journal of Political Economy* 80 (May/June): 435–63.
Averitt, R. T. 1968. *The Dual Economy.* New York: Norton.
Beck, E. M., P. M. Horan, and C. M. Tolbert II. 1978. "Stratification in a Dual Economy: A Sectoral Model of Earnings Determination." *American Sociological Review* 43 (October): 704–20.
———. 1980. "Industrial Segmentation and Labor Market Discrimination." *Social Problems* 2 (December): 113–30.
Bibb, R., and W. Form. 1977. "The Effects of Industrial, Occupational, and Sex Stratification on Wages in Blue-Collar Markets." *Social Forces* 55 (June): 974–96.
Bishop, Y. M. M., S. E. Fienberg, and P. W. Holland. 1975. *Discrete Multivariate Analysis: Theory and Practice.* Cambridge, Mass.: MIT Press.
Bluestone, B. 1970. "The Tripartite Economy: Labor Markets and the Working Poor." *Poverty and Human Resources* 5 (July/August): 15–35.
Daymont, T. N., and R. L. Kaufman. 1979. "Measuring Industrial Variation in Racial Discrimination Using Log-linear Models." *Social Science Research* 8 (1): 41–62.
Doeringer, P., and M. J. Piore. 1971. *Internal Labor Markets and Manpower Analysis.* Lexington, Mass.: Heath.
Duncan, O. D. 1968. "Inheritance of Poverty or Inheritance of Race?" Pp. 85–110 in *On Understanding Poverty,* edited by D. P. Moynihan. New York: Basic.
Edwards, R. C. 1979. *Contested Terrain.* New York: Basic.
Featherman, D. L., and R. M. Hauser. 1976. "Changes in the Socioeconomic Stratification of the Races, 1962–1973." *American Journal of Sociology* 82 (November): 621–51.
Franklin, R. S., and S. Resnik. 1973. *The Political Economy of Racism.* New York: Holt, Rinehart & Winston.
Gordon, D. M. 1972. *Theories of Poverty and Underemployment.* Lexington, Mass.: Heath.
Hansen, W. L., B. A. Weisbrod, and W. J. Scanlon. 1970. "Schooling and the Earnings of Low Achievers." *American Economic Review* 60 (June): 409–19.
Hill, R. C. 1974. "Unionization and Racial Income Inequality in the Metropolis." *American Sociological Review* 39 (August): 507–22.
Hodson, R. 1978. "Labor in the Monopoly, Competitive, and State Sectors of Production." *Politics and Society* 8 (3–4): 429–80.
———. 1983. *Corporate Economic Structure and Workers' Earnings.* New York: Academic Press.
Hodson, R., and R. L. Kaufman. 1982. "Economic Dualism: A Critical Review." *American Sociological Review* 47 (December): 727–39.
Kalleberg, A. L., M. Wallace, and R. P. Althauser. 1981. "Economic Segmentation, Worker Power, and Income Inequality." *American Journal of Sociology* 87 (November): 651–83.
Kaufman, R. L. 1981. *Racial Discrimination and Labor Market Segmentation.* Ph.D. dissertation, University of Wisconsin—Madison.
Kaufman, R. L., and T. N. Daymont. 1981. "Racial Discrimination and the Social Organization of Industries." *Social Science Research* 10 (3): 225–55.
Kaufman, R. L., R. Hodson, and N. D. Fligstein. 1981. "Defrocking Dualism: A New Approach to Defining Industrial Sectors." *Social Science Research* 10 (1): 1–31.

Kluegel, J. R. 1978. "The Causes and Costs of Racial Exclusion from Job Authority." *American Sociological Review* 43 (June): 285–301.

Marshall, R. 1974. "The Economics of Racial Discrimination: A Survey." *Journal of Economic Literature* 12 (September): 849–71.

Masters, S. H. 1975. *Black-White Income Differentials: Empirical Studies and Policy Implications*. New York: Academic Press.

Olneck, M. R. 1976. "The Effects of Education on Occupational Status and Earnings." Institute for Research on Poverty Discussion Paper no. 358–76. Madison: University of Wisconsin—Madison.

Schervish, P. G. 1981. "The Structure of Employment and the Structure of Unemployment." Pp. 153–86 in *Sociological Perspectives on Labor Markets*, edited by I. Berg. New York: Academic Press.

———. 1983. *Vulnerability and Power in Market Relations: The Structural Determinants of Unemployment*. New York: Academic Press.

Shepherd, W. G. 1970. *Market Power and Economic Welfare*. New York: Random House.

Siegel, J. S. 1974. "Estimates of Coverage of the Population by Sex, Race, and Age in the 1970 Census." *Demography* 11 (February): 1–23.

Smith, J. D., and F. R. Welch. 1977. "Black-White Male Wage Ratios: 1960–1970." *American Economic Review* 67 (June): 323–38.

Snyder, D., and P. M. Hudis. 1976. "Occupational Income and the Effects of Minority Competition and Segregation: A Reanalysis and Some New Evidence." *American Sociological Review* 41 (April): 209–34.

Spilerman, S. 1977. "Careers, Labor Market Structure, and Socioeconomic Achievement." *American Journal of Sociology* 83 (November): 551–93.

Stolzenberg, R. M. 1975. "Education, Occupation, and Wage Differences between White and Black Men." *American Journal of Sociology* 81 (September): 299–323.

Thurow, L. C. 1975. *Generating Inequality*. New York: Basic.

Tolbert, C. M., P. M. Horan, and E. M. Beck. 1980. "The Structure of Economic Segmentation: A Dual Economy Approach." *American Journal of Sociology* 85 (March): 1095–1116.

U.S. Bureau of the Census. 1972. *Public Use Samples of Basic Records from the 1970 Census: Descriptive and Technical Documentation*. Washington, D.C.: Government Printing Office.

Wallace, M., and A. L. Kalleberg. 1981. "Economic Organization, Occupations, and Labor Force Consequences: Toward a Specification of Dual Economy Theory." Pp. 77–117 in *Sociological Perspectives on Labor Markets*, edited by I. Berg. New York: Academic Press.

Winsborough, H. H., and P. J. Dickinson. 1971. "Components of Negro-White Income Differentials." Pp. 6–8 in the *Proceedings of the Social Statistics Section, Meeting of the American Statistical Association*. Fort Collins, Colo.: American Statistical Association.

Wright, E. O. 1978. "Race, Class, and Income Inequality." *American Journal of Sociology* 83 (May): 1368–97.

Zucker, L. G., and C. Rosenstein. 1981. "Taxonomies of Institutional Structure: Dual Economy Reconsidered." *American Sociological Review* 46 (December): 869–84.

Sex and Earnings in Industrial Society: A Nine-Nation Comparison[1]

Donald J. Treiman
University of California, Los Angeles

Patricia A. Roos
State University of New York at Stony Brook

A substantial difference in average earnings between men and women employed full-time is documented for each of nine industrial nations, and several hypothesized explanations for the earnings gap are explored: a human capital hypothesis—women earn less because they have less education and experience; a dual career hypothesis—women earn less because they adjust their work behavior to meet the demands of family obligations; and an occupational segregation hypothesis—women earn less because they are concentrated in low-level jobs. None of these hypotheses receives much support in any country, leaving open the possibility that the earnings differences are due to deeply entrenched institutional arrangements that limit women's opportunities and achievements.

Throughout the industrialized world, women earn substantially less than men for the work they do. Although fully systematic data are lacking, various estimates of the relationship between male and female average earnings have been published for a number of countries. In the United States, for example, the median annual earnings of women aged 14 and older were 43% of those of men in 1978 (U.S. Department of Labor 1980, table 53). Similarly, in Japan women's earnings averaged 50% of those of men in 1972 (Takahashi 1976). In some Eastern European countries, the gap is smaller—in the USSR in 1963, women earned 65% as much

as men (Swafford 1978, p. 661), and in Poland in 1973, 68% as much (computed from Haavio-Mannila and Sokolowska [1978], p. 204). Even when only full-time workers are considered, women's earnings on average fall far short of those of men: in the United States the ratio is 57%, a ratio that remained essentially constant from 1960 through 1978 (U.S. Bureau of the Census 1980). Similarly, in Canada in 1972, full-time women workers earned 62% as much as male employees (Boyd and Humphreys 1979). As we shall see below, in most Western European countries the ratios tend to be somewhat higher, up to 74% in the Federal Republic of Germany. When hourly pay rates within particular occupational sectors are considered, the ratios are still higher, averaging 74% for manufacturing jobs in 10 industrialized European nations in 1977 (computed from International Labour Office [1977], table 19A).

How might we account for this gap? A variety of explanations have been proposed, which we review below. To date, most research on this topic has been restricted to the United States, leaving it unclear to what extent the degree and pattern of gender differences in income reflect idiosyncratic historical developments within this country and to what extent such differences reflect fundamental features of the organization of work and family life across industrial societies. While the evidence is fairly clear, if somewhat unsystematic, that across industrial societies men on average earn more than women, there has not as yet been any serious attempt to test the generality of alternative explanations for these differences on a comparative basis.

This paper represents a first attempt to address this issue systematically. We compare the incomes of men and women in nine industrial societies and then, insofar as data permit, attempt to account for the observed average differences by testing a variety of hypotheses. Because of the paucity of comparable cross-national data on gender differences in income and their determinants, we regard the descriptive aspect of our effort as particularly important and thus present the basic data underlying our analysis rather more fully than might otherwise have been necessary. Before turning to the presentation of the data and analysis, however, we review various explanations of the gap between the average earnings of men and women.

EXPLANATIONS FOR THE MALE-FEMALE EARNINGS DIFFERENTIAL

There are four major classes of explanations for the earnings differential between men and women: (1) women's productivity is lower than that of men because they have less accumulated human capital, (2) women's family responsibilities affect their earning capacity, (3) women do different work than men and the work women do pays less than the work men

613

do, and (4) women are subject to discriminatory market mechanisms other than occupational segregation that affect their earnings relative to men's.[2] We will be able to provide explicit, albeit partial, tests of the first three explanations; the validity of the fourth must be inferred indirectly from the results of the analysis.

Human Capital

Prominent among explanations of the gender difference in earnings is one derived from the human capital theory of economics.[3] Proponents of this theory argue that superior qualifications enhance the productivity of workers and that pay differences reflect differential productivity; those workers with the best qualifications will thus be the most highly paid. Because direct measures of productivity are seldom available, the two-step path from qualifications to productivity to earnings is virtually never empirically investigated. Instead, worker characteristics assumed to affect productivity, such as amount of formal education, on-the-job training, and labor force experience, are used to predict earnings. Thus, according to human capital theory, insofar as women have less of these attributes than men, they should earn less. In a general way, the American evidence bears this out—formal schooling, on-the-job training, and experience have been shown to be positively related to the earnings of individuals and to account partially for the gender difference in average earnings (Mincer and Polachek 1974, but see Sandell and Shapiro 1978; Kohen 1977; Treiman and Terrell 1975b, p. 196). Whether the link between education and experience, on the one hand, and earnings, on the other, is, in fact, explained by productivity differences is problematic and a matter of considerable controversy. We are unable to enter that debate here and restrict ourselves to investigating what part of the earnings gap between men and women in other industrialized countries can be attributed to gender differences in education and labor force experience.

Dual Career Responsibilities

While current family demands, or expectations about future family responsibilities, may affect decisions to invest in educational or on-the-job training (Mincer and Polachek 1974) and decisions about what occupation to enter, there are other more direct effects of women's family responsibilities on their earnings. Specifically, family responsibilities may affect

[2] For an extended discussion of these issues see Roos (1981a).

[3] For additional detail on human capital theory and its relation to explanations of gender differences in earnings, see Kahne and Kohen (1975), Blau and Jusenius (1976), Treiman and Hartmann (1981, chaps. 2 and 3), and England (1982).

a woman's choice of jobs within occupations and the time and energy she can devote to career advancement. Whether or not they work outside the home, women in industrial societies bear the primary responsibility for time-consuming home and family chores (Szalai 1972; Walker and Woods 1976). Because in industrial societies most market work is located outside the home, women who work must organize their family and work lives to balance these often conflicting demands on their time. For many married women, the existence of dual demands may require choosing jobs that minimize conflict with family responsibilities (e.g., jobs located near their home or those not requiring substantial overtime or travel) rather than those that maximize pay. Moreover, women may interrupt their labor force participation when working conflicts with family responsibilities and hence lose seniority rights and, as a result, find themselves in an inferior bargaining position when they return to the work force. Finally, even when married women remain in the labor force, their family responsibilities may make it difficult to perform on the job in such a way as to increase the probability and speed of pay increases and promotions.

Occupational Segregation

It is well-known that the American occupational structure is highly segregated by sex—indeed, much more highly segregated by sex than by race. The degree of segregation shows little sign of lessening (Gross 1968; Treiman and Terrell 1975c), and the small changes that have occurred can be attributed to the movement of men into traditionally female jobs and not to any change in the distribution of women across the occupational structure (Blau and Hendricks 1979).

Men and women tend to be concentrated in substantially different occupational sectors, with women overrepresented in clerical and service jobs and underrepresented in managerial, craft, and laboring jobs (Oppenheimer 1970). In particular, women tend to be concentrated in occupations that pay poorly relative to the average educational achievement of incumbents (Oppenheimer 1968; Treiman and Terrell 1975c). Even within occupational sectors, men and women are concentrated in different jobs, and the jobs women hold tend to pay less. For example, male physicians tend to be mainly in private practice, while female physicians work mainly in salaried positions (Kosa and Coker 1971, p. 199). Similarly, male professors are more likely to teach at major universities than female professors (Astin and Bayer 1975, p. 375). Even when job tasks are relatively homogeneous within occupations (e.g., accounting clerk, order clerk, tabulating machine operator), women tend to be paid less than men in the same occupation. Several studies have found that the higher the proportion of women doing a particular job within a firm, the

615

less that job pays relative to the same job in other firms (McNulty 1967; Buckley 1971; Blau 1977).

We are not concerned in the present analysis with accounting for occupational segregation by sex, which may result from a host of factors, but only with the consequences of occupational segregation for pay differentials by sex. In considering this question, it is useful to distinguish between occupational pay differentials based on job content and those based on other factors. It is widely accepted, both as a normative prescription and as an empirical description (Treiman 1977, chap. 1), that jobs requiring the highest degree of skill and responsibility will be most highly rewarded. Hence, to the extent that men do work that is more demanding than work done by women, we would expect them to be paid more on average. In addition, there are other factors that create differential pay rates—the scarcity of labor, sometimes manipulated by restricting the supply of eligible workers; the profitability of firms or economic sectors; the power of unions; legislated rates; tradition; and discrimination. It has been suggested that such factors are important in explaining the male-female earnings differential in the United States (Treiman and Hartmann 1981).

Discrimination

Finally, the social characteristics of individual workers may also affect their earnings. Sometimes pay differentials based solely on worker attributes (without regard to performance) are regarded as legitimate and sometimes they are regarded as discriminatory. In the United States, until 1963 it was legal, and widely accepted as legitimate, to pay women less than men doing exactly the same work. It is still regarded as appropriate to base pay differences on seniority.[4] Hence, insofar as male workers tend to have more seniority because of their more continuous work history, we would expect them to receive higher pay than female workers. In industrial countries it is no longer regarded as legitimate to base pay differentials on gender. Most industrial nations either have their own equal pay legislation or support article 1 of International Labour Office (ILO) Convention 100, which calls for equal pay for "work of equal value."[5] Still, progress toward equal pay has been slow, as the large observed gender disparity attests (see also Wallace 1976; Loeb, Ferber, and Lowry 1978; O'Kelley 1979).

There is every reason to suspect, therefore, that even when differences in the type of work performed, the qualifications of workers (including

[4] A seniority difference is one of the bases on which a sex differential in earnings can be justified according to the 1963 Equal Pay Act (29 U.S. Code S206[d]).

[5] Those countries belonging to the European Economic Community (EEC) are also subject to the provisions of the Treaty of Rome that call for "equal pay for equal work."

the amount of labor force experience), and the extent of family obligations are all taken into account, a substantial difference in the earnings of male and female workers will remain.

DATA AND METHODS

Data

This analysis compares the earnings of male and female workers in nine industrialized countries: Austria, Denmark, Finland, Germany (Federal Republic), Israel, the Netherlands, Norway, Sweden, and the United States. These countries are a subset of the approximately 30 countries for which representative sample survey data have been obtained in conjunction with an ongoing comparative study of social mobility (Treiman and Kelley 1974). This analysis, like that of the larger project, involves reanalysis of data originally collected for other purposes. Considerable effort was thus required to standardize the variables and samples in order to permit valid cross-national comparisons. These efforts will be described where relevant below (see Roos [1981a] for a more complete description of the data sets used and variable transformations made).

The sample of countries was restricted to industrial societies in order to minimize the problem of identifying the economically active population. Serious problems of noncomparability in measured rates of female labor force participation occur in developing countries since in those countries a large fraction of women are unpaid family workers and such workers are sometimes included in and sometimes excluded from the "economically active" population (Durand 1975, p. 11). Cross-national comparisons of nonindustrial and industrial societies would thus be seriously hampered by a lack of comparability in employment data.

In all cases the data are based on representative national probability samples of the adult population (sources of the data sets are given in the notes to table 1). Because of variability in the age groups included in the samples as well as variability in the age at which labor force participation normally begins, we have restricted the bulk of our analysis to the most economically active portion of the population, persons aged 20–64.

We have also restricted the analysis of earnings differences to the full-time employed population in each country. While this decision reduces substantially the number of female workers, the absence of data on hours worked by part-time workers gave us little choice since it is well-known that the earnings of part-time female workers are heavily dependent on the amount of time spent working (Treiman and Terrell 1975b).[6]

---

[6] There may, of course, be differences in hours worked even in the "full-time" labor force, but the variance will be reduced substantially relative to the total labor force. Mcreover, it is not evident whether hours worked matters among jobs that are defined as full-time.

Table 1 gives the distribution of employment status (full-time, part-time, or not in labor force) by sex for each country. Appendix A presents an elaboration of table 1, providing the percentage employed full-time by age and sex. As can be seen from the table, there is substantial variability across countries in the percentage of women working and their propensity to work full-time. In addition, the results in Appendix A suggest that there is cross-cultural variability in women's age patterns of labor force participation. These cross-cultural differences may in turn affect the kinds of occupations men and women enter and/or the earnings they receive. While we refer to these patterns at various points in discussing our results, a detailed analysis of why such differences exist, and what effect they have on occupational and income attainment, is beyond the scope of this paper. Readers interested in exploring these issues further are referred to the basic data presented in Appendixes A and B and in Roos (1981a).

Table 1 also serves as a good illustration of the difficulties inherent in the secondary analysis of data for comparative purposes and suggests the

TABLE 1

LABOR FORCE STATUS, BY SEX, for NINE INDUSTRIALIZED COUNTRIES (%)

Country	Employed Full-Time	Employed Part-Time	Not in Labor Force	Total	N
Females:					
Austria................	31.5	9.1	59.5	100.1	867
Denmark.............	34.1	22.3	43.6	100.0	560
Finland	55.0	16.3	28.7	100.0	509
Federal Republic of					
·Germany...........	18.7	15.4	65.9	100.0	1,294
Israel	16.0	13.6	70.4	100.0	15,013
Netherlands	25.0	9.3	65.7	100.0	471
Norway..............	21.4	21.9	56.7	100.0	462
Sweden	28.0	37.7	34.3	100.0	554
United States	30.3	11.9	57.8	100.0	3,067
Males:					
Austria................	79.5	.5	20.1	100.1	683
Denmark......·......	83.0	7.0	10.0	100.0	458
Finland	70.9	14.5	14.7	100.1	484
Federal Republic of					
Germany...........	72.9	3.3	23.8	100.0	889
Israel	54.4	12.9	32.6	99.9	13,220
Netherlands	68.5	1.8	29.7	100.0	663
Norway..............	77.5	10.1	12.4	100.0	516
Sweden	77.8	14.3	7.9	100.0	468
United States	60.8	8.8	30.3	99.9	2,837

SOURCES.—Austria, Netherlands (Political Action: An Eight-Nation Study 1973–76 [1979]); Denmark, Finland, Norway, Sweden (Scandinavian Welfare Survey 1972 [1977]); Federal Republic of Germany (Zumabus 1976 [1976]); Israel (Israel Labor Mobility Survey 1974 [1977]); United States (NORC General Social Surveys 1974–77 [1977]).

NOTE.—See n. 7 for the definition of the labor force status variable.

necessity of exercising caution in the interpretation of results. Because the variables from which we constructed our measure of employment status differ from country to country, the construction of a single variable that would be comparable across countries required a number of (sometimes arbitrary) assumptions.[7]

Analytic Strategy

Our basic approach is to estimate regression equations separately for males and females in each country, predicting income from a variety of variables identified in our discussion. In a second step, we use a regression standardization procedure to disaggregate the male-female income gap into a component due to differences in male and female means on the independent variables included in the regression analyses and a component due to differences in the rate of return to these variables.

Variables

Four determinants of earnings are analyzed—educational attainment, experience, marital status, and occupational position. We discuss these in turn, showing their relationship to the explanations for the earnings gap reviewed in the preceding section.

   Educational attainment.—Education and experience are used as (imperfect) indicators of human capital investments. We measure educational attainment by years of school completed. While amount of schooling does not fully capture variation in the quality of education in countries that have complex educational systems (see Treiman and Terrell [1975a] for a detailed discussion regarding the measurement of educational attainment in Great Britain and the United States), we restrict ourselves to this variable because we lack more detailed data for most of the countries considered here. Moreover, our main focus is on the comparison of males and females within countries and not on comparisons between countries,

[7] The labor force status variables were defined as follows: Austria, the Netherlands ("full-time" = currently working full-time; "part-time" = currently working part-time; "not in labor force" = former workers, housewives, retired and disabled, those who have never worked, those in school, and those in the military); Denmark, Finland, Norway, Sweden ("full-time" = full-time nine or more months during 1971; "part-time" = part-time work all year or full-time work less than nine months during 1971; "not in labor force" = those who did not work at all during 1971, e.g., students, pensioners, housewives, persons under institutional care); Federal Republic of Germany ("full-time" = gainful employment full day; "part-time" = gainful employment half day or incidentally gainfully employed; "not in labor force" = not gainfully employed); Israel ("full-time" = employed full-time last week; "part-time" = employed part-time last week for at least 15 hours a week or temporarily absent, worked during year; "not in labor force" = unemployed, worked less than 15 hours a week, or not in civilian labor force, e.g., army, unpaid family workers); the United States ("full-time" = full-time last week; "part-time" = part-time last week or temporarily absent; "not in labor force" = unemployed, retired, in school, or housewife).

and it is unlikely that years of school completed is a more valid measure of educational attainment for one sex than for the other.[8]

. We expect education to have a strong positive effect on the earnings of both men and women. The earnings gap between males and females thus could simply reflect differences in achieved education. Although in the United States men and women tend on average to have equal amounts of schooling (Treiman and Terrell 1975b), schooling differences may account for gender differentials in earnings in other industrial societies. Insofar as this is so, the earnings gap could be viewed as legitimate. However, if males and females have different rates of return on their investments in education, the human capital hypothesis is called into question.

Experience.—Because the data sets available to us do not have any direct measures of experience or seniority, we include a measure of potential experience, estimated by the conventional formula:

potential experience = age − years of schooling − 6.

Although this measure provides a reasonably good estimate of the total years of experience for men, since most men work full-time continuously after completing their schooling, it does not provide an equally adequate estimate of women's experience. While never-married women tend to have continuous attachment to the labor force (Treiman and Terrell 1975b), the participation of married women tends to be part-time and intermittent.

The bias introduced by using this measure may not, however, be as great as is often assumed. Reeve Vanneman, using data from the 1972–73 Quality of Employment Survey (1975), has shown (personal communication, 1981) that although the "potential experience" variable substantially overestimates actual labor force experience for women (and not for men), regression equations using actual and potential experience produce substantially similar results. In particular, the differences between equations estimated for males and females using either measure of experience are much larger than the differences between the equations for each sex using the alternative measures.[9]

[8] It is an open question whether, for a given number of years of school completed, quality of education is more differentiated for males or for females and whether such differences as exist are consistent across countries. With respect to postsecondary education in the United States, curriculum is undoubtedly important—few would dispute that a B.A. in engineering is worth more financially than a B.A. in history or that an A.A. (two-year) certificate in medical technology is worth more than an A.A. in marketing. But it is by no means obvious whether the courses of study dominated by men are more differentiated than those dominated by women, or vice versa.

[9] For national samples of males and females employed full-time, the pertinent regression estimates are:

For males:
$\hat{Y} = -3,204 + 482(E) + 1,827(M) + 57.3(S) + 415(X) - 6.37(X^2)$, $R^2 = .232$;
$\hat{Y} = -3,938 + 499(E) + 1,615(M) + 61.2(S) + 451(X') - 6.75(X'^2)$, $R^2 = .242$.

Because it is well-established that the cross-sectional relationship between experience and earnings is curvilinear, with older workers experiencing a reduction in earnings relative to those in midcareer, we include a squared term for experience in our estimation equations. Specifically, we include the variable $(X - k)^2$, where X is our potential experience variable and the constant, k, is chosen to make the squared term orthogonal to the experience variable.[10] The advantage of doing this is that it makes it possible to interpret the effects of experience and the squared term independently. A curvilinear relation between experience and earnings may arise in a number of ways: for some occupations, productivity may fall with age after a peak is reached, owing to declining physical capabilities; in others, skills acquired as part of formal education may depreciate to a point that more than offsets increases in productivity associated with experience. Alternatively, there may be no decline in earnings within any particular cohort; the observed curvilinear cross-sectional relationship may simply reflect cohort differences in earnings, each cohort starting and ending with higher earnings than previous cohorts (for an example of this process as it pertains to occupational status see Jackson, Thompson, and Treiman [1979]). While we will be unable to decide among these possibilities here, the inclusion of the squared term for experience is necessary to specify the estimation equation properly.

Marital status.—Marital status is a dummy variable, with those who have never married assigned a value of 1 and others a value of 0. Its utility in the present analysis is in testing the hypothesis that women's family and work obligations affect earnings negatively. Insofar as the dual burden faced by working women with families results in occupational decisions or performances that reduce earnings, never-married women

---

For females:
$$\hat{Y} = -2,083 + 353(E) + 50.8(M) + 47.3(S) + 211(X) - 3.54(X^2), \quad R^2 = .334;$$
$$\hat{Y} = -2,539 + 370(E) + 240(M) + 52.5(S) + 184(X') - 2.71(X'^2), \quad R^2 = .327.$$

Here Y = annual income, E = years of school completed, M = 0 if never married and 1 if ever married, S = the Duncan SEI score for current occupation, X = the actual number of years worked since age 16, and X' = potential number of years worked, estimated as age minus years of schooling minus 4. The means for X and X' are 20.8 and 22.5 for males and 15.9 and 22.3 for females. Although the potential experience variable overestimates actual labor force experience much more severely for females than for males, the impact on the regression coefficients is less substantial. Even in the female equation, only one coefficient changes substantially, the statistically insignificant coefficient for marital status. More important, the conclusions regarding gender differences in the determinants of income that one would draw from a comparison of the equations using the estimated experience variable are the same as the conclusions one would draw from a comparison of the equations using the actual experience variable.

[10] Kent Smith has shown (personal communication, 1981) that choosing k equal to one-half the slope coefficient associated with the zero-order regression of X^2 on X will make X and $(X - k)^2$ orthogonal. This is a special case of the procedure discussed in Smith and Sasaki (1979).

should earn more than ever-married women, net of other factors. The distinction has been drawn between never- and ever-married women, rather than between currently married and not currently married women, because the effect of a dual career is likely to persist even if a marriage is dissolved—career disruptions may have a permanent negative effect on earnings even when a full-time career is resumed. Moreover, children often remain from a former marriage, and traditionally they are the mother's rather than the father's responsibility. Empirical evidence confirming the similarity between currently and formerly married women in their earnings attainment and the dissimilarity between never- and ever-married women is provided by Treiman and Terrell (1975b, p. 189).[11]

Whereas marriage may be regarded as a burden for working women, it is likely to be an advantage for working men. The fact that responsibility for household maintenance largely falls to the wife (Szalai 1972; Walker and Woods 1976) leaves the husband free to pursue his career single-mindedly. Of course, the added financial burden of maintaining a family may lead men, more than women, to seek jobs that maximize their incomes (see Oppenheimer 1974).

Occupation.—Our data permit two measurements of occupational position: prestige and a classification of occupations into the seven major categories of the *International Standard Classification of Occupations* (International Labour Office 1969).

Consider prestige first. Recall the suggestion that occupational pay differentials can be divided into those based on job content and those based on other factors. In the present context we can regard the prestige of occupations as a summary indicator of those aspects of job content that are the major bases of compensation. This view derives from Treiman's (1977, chap. 1) theoretical work that identifies as major determinants of prestige various aspects of power: economic control, authority, and expertise. Interestingly, these conform closely to the major compensable factors in formal job evaluation schemes. Although skill, effort, responsibility, and working-conditions factors are generally included in such schemes, invariably major weight is accorded to skill and responsibility and only minor weight is accorded to effort and working conditions (Treiman 1979). Prestige is measured by the Standard International Occupational Prestige Scale (Treiman 1977, chap. 8) by assigning prestige scores to the detailed occupational categories into which data were coded in each of the nine countries. Such assignments have been shown to be highly reliable, and the resulting prestige scores have been shown to have high validity (Treiman 1977, chap. 9).

---

[11] Interestingly, available evidence for the United States indicates that the presence or number of children has virtually no impact on the earnings of ever-married women (Treiman and Terrell 1975b, pp. 190, 195). Unfortunately, most of our samples do not contain the data necessary to test this possibility in other countries.

Of course, job content, as measured by prestige, is not the only basis for occupational differentials in pay rates (Bridges and Berk 1974; McLaughlin 1975; Talbert and Bose 1977). In fact, the use of prestige alone would be inappropriate. It is well-known that men and women tend to work in substantially different kinds of jobs even though on average they work in jobs of equivalent prestige (England 1979). The use of prestige as a measure of occupational attainment is limited in analyses of sex differences in earnings precisely because it fails to distinguish between the kinds of jobs in which men and women work (Roos 1981b). Because we expect occupational sectors to differ in their level of compensation as a consequence of differences in their mode of organization (degree of unionization, effective control of the supply of labor, etc.), and because the labor force is so highly segregated by sex, sectoral differences in compensation levels could themselves produce substantial male-female earnings differentials. While the limitations of our data do not permit us a very fine-grained consideration of these various effects, we can make a first approximation by distinguishing among the seven major groups of the *International Standard Classification of Occupations* (International Labour Office 1969). While it would have been preferable to use a more detailed classification, such as the 14-category classification proposed by Treiman (1977, chap. 9), the small samples available to us precluded this. However, by scoring each respondent with both the prestige of his or her occupation and a dummy variable representing the major group in which the occupation is located, we are able to capture both sectoral differences in income levels net of prestige and prestige differences in income within major occupational sectors.

Income.—Our dependent variable is income. As the reader will note, we have been using "income" and "earnings" somewhat interchangeably. Our data are sometimes earnings (Austria, Germany, Israel, the Netherlands, and the United States) and sometimes income (Denmark, Finland, Norway, and Sweden). Fortunately, for practical purposes the distinction is unimportant. Except for the extreme ends of the distribution, the overwhelming bulk of income derives from earnings, at least in the United States. Only the very poor, who subsist largely on transfer payments, and the very rich, who have substantial returns from investments, have enough outside income to affect the analysis.[12] In sample surveys of the kind we have available, these extreme groups are likely to be represented by only a few cases, if at all.

The exact information elicited on income varies from country to country, including net or gross monthly earnings, gross annual earnings, and gross annual income (see the note to table 4). After converting categorical

[12] In the United States in 1972, 84% of gross income derived from wages and salaries and over 90% from wages and salaries or profits from farms, partnerships, and businesses (computed from U.S. Internal Revenue Service 1974, table 1.4).

variables to an interval scale by assigning each category its midpoint and estimating the midpoint of the open-ended upper categories by means of a Pareto transformation (Miller 1966), we converted each income variable into its natural logarithm. Doing this effectively standardizes the income variables across the countries: when the dependent variable is the natural log of income, the metric regression coefficients can be interpreted as indicating approximately the proportional increase in income that would be expected from a one-unit increase in the associated independent variable.[13] A further justification for logging income, in addition to its interpretive convenience, is that the distribution of earnings tends to be lognormal.

ANALYSIS

The analysis will proceed in four steps: first, we examine gender differences in occupational composition in all nine countries; second, we examine gender differences in income; third, we estimate a regression model of income determination separately for males and females; and, finally, we use the results of our regression analyses to decompose the gender differences in income into a number of factors, as a way of testing alternative explanations for the observed income gap.

Occupational Composition

It is evident that in industrial societies men and women do different work. Table 2 presents the distribution by sex of the entire labor force (both full- and part-time) over the 14 categories of Treiman's (1977, pp. 203–8) International Occupational Classification.[14] While there is some variability in the degree of occupational segregation of the labor force by sex, with indexes of dissimilarity ranging from 38 in Austria to 60 in Sweden, the general pattern is one of very substantial segregation in all nine countries.

---

[13] This approximation holds for values of b (the net regression coefficient) up to about .2. For higher values of b the proportional increase is underestimated by successively larger amounts: $b = .2$ implies an increase of 22%, $b = .3$ implies an increase of 35%, $b = .4$ implies an increase of 49%, and $b = .7$ implies an increase of 101%.

[14] This classification, which was formed by dividing the seven-category *International Standard Classification of Occupations* (International Labour Office 1969) on the basis of prestige, effectively captures major variations in the organization of work, e.g., dividing "professional, technical, and related occupations" into high-prestige professions (which correspond to the classical liberal professions) and low-prestige professions (which correspond to the technical and semiprofessions). Such a distinction is particularly important in comparing the earnings of men and women, given the greater likelihood of women's concentration in low-prestige professions (e.g., primary school teaching, nursing, and librarianship). Similar distinctions are drawn throughout the labor force. In tables 2 and 3 the categories are arranged in order of their average prestige.

In order to see more clearly how women are distributed throughout the labor force, that is, in which sectors they are overrepresented and in which they are underrepresented, we turn to table 3. This table presents the percentage female in each occupational category expressed as a deviation from the percentage female in the entire labor force. This standardization adjusts for the fact that in some countries women are much more likely to work than in others. For example, in Norway the percentage of high-prestige professionals who are female is 6% lower than the percentage female in the labor force as a whole. Similarly, in Sweden the percentage of high-prestige professionals who are female is 8% lower than the percentage female in the Swedish labor force. Hence, we conclude that women are about equally underrepresented among high-prestige professionals in Norway and Sweden, even though 21% (= 27% − 6%) of high-prestige professionals in Norway are women and 35% (= 43% − 8%) of high-prestige professionals in Sweden are women.[15]

Inspection of the table shows the degree of similarity across countries to be striking. In nearly all countries, women are substantially overrepresented in high-prestige clerical occupations and in low-prestige sales and service occupations; slightly overrepresented in low-prestige professional occupations, high-prestige service occupations, and low-prestige clerical occupations; slightly underrepresented in high-prestige professional, sales, and agricultural occupations and in low-prestige production occupations; and substantially underrepresented in administrative occupations, in high- and medium-prestige production occupations, and in low-prestige agricultural occupations. Of 126 possible comparisons, only nine (or 7%) have a sign opposite the major trend for the row (these are indicated by boldface), and in almost all cases these deviations are very small.[16] Clearly, there is a characteristic pattern of men's work and wom-

[15] Our samples were adjusted to account for the fact that in many of the countries, working men and women were sampled at different rates (owing mainly to differential response rates). Without such an adjustment, the computation of percentage female in each occupational category of table 3 would have produced incorrect estimates. To correct this bias, we weighted the data so that the total percentage female in each sample equals the percentage female in the labor force reported in the 1977 *Yearbook of Labour Statistics* (International Labour Office 1977), except in the case of the United States, for which a 1976 Current Population Survey estimate was used (U.S. Bureau of the Census 1977). Care was taken to select a census date that was close to the date of the survey. This correction assumes that within each country the differential response by sex is the same for each occupational category. Since the remainder of our analysis involves separate analyses by sex, the weighting procedure affects only the results presented in table 3.

[16] There is only one truly anomalous statistic in the entire set—the substantial overrepresentation of German women among low-prestige production workers. We suspect that this may reflect a peculiarity of the German sample survey design. The German survey organization that did the field work, ZUMA, includes only German nationals in its samples. Hence, the *Gastarbeiter* ("guest worker") population, which formed 11% of the total labor force in 1972 (Federal Republic of Germany, Federal Institute for Population Research 1974, p. 32), is excluded. This population, which is largely male, is concentrated in the lowest-

en's work in industrial societies. While it is beyond the scope of the present paper to attempt to account for this pattern, its existence does suggest that in our investigation of the earnings differential we are concerned with very fundamental features of the organization of work in industrial societies.

Income

Table 4 simply confirms what the material presented in the introductory section suggested—even among full-time workers, women earn substantially less than men. Moreover, the gender gap in earnings is greatest in the United States: during the period 1974–77, women who worked full-time earned an average of 51% of what men earned per year.[17] At the other extreme, the gap is smallest in the Federal Republic of Germany, where women earned an average of 74% of what men earned. The nine-country average is 64%—in these countries the average woman working full-time earned less than two-thirds of what her male co-workers earned. We now turn to discussion of why this very large gender discrepancy exists.

Determinants of Earnings

Our first step in accounting for the average earnings gap between male and female workers is to predict earnings separately for full-time workers of each sex within each country. We thus estimated 18 equations (two sexes by nine countries) of the form:

$$\hat{\ln Y} = a + b(E) + c(X) + d(X'^2) + e(S) + f(P) + \Sigma g_i D_i, \quad (1)$$

status jobs. Hence, it may well be the case that in Germany, as elsewhere, low-status production work is done mainly by men but by *foreign* rather than *German* men. This possibility is supported by the data in table 2 showing that the proportion of the male labor force in low-prestige production occupations is lower in Germany than in any of the other eight countries.

[17] The U.S. ratio may well be a lower-bound estimate. The discrepancy between our ratio of .51 and that of .59, the average of the ratios reported in the Current Population Surveys (CPS) for the same years (computed from U.S. Department of Labor 1979, table 1) can perhaps be accounted for by the difference in the definition of "full-time" work in the NORC surveys we used and the CPS. The NORC surveys asked only whether the individual worked full-time in the week preceding the survey. The CPS asked how many hours the respondent worked in the reference week and also how many weeks the respondent worked in the past year; full-time year-round workers are defined as those who worked 50–52 weeks in the past year and 35 or more hours in the past week. Since the CPS definition is more restrictive, the ratio of female to male earnings is bound to be somewhat higher in the CPS data. For example, primary school teachers—an occupational category with more female than male incumbents—would be counted as full-time workers in the NORC data but not in the CPS data.

where $\ln Y$ = natural log of income; E = years of school completed; X = potential experience, estimated by age − years of schooling − 6; X' = $X - k$, with k chosen to make X and X'^2 orthogonal (see the discussion above); S = 1 for those who have never married, and 0 otherwise; P = the prestige of the respondent's occupation; and each D_i = 1 for occupations in the ith major group of the *International Standard Classification of Occupations*, and 0 otherwise. The first category, professional and technical occupations, is omitted from the equation to avoid a linear dependency. The means, standard deviations, and correlations among the included variables are given in Appendix B, which also includes the linear form of the income variable and the omitted dummy variable for professional and technical occupations. The regression estimates are given in table 5. We discuss the coefficients of each variable in the order in which we introduced them above.

Education.—While education has a positive effect on income for men, the same is not true for women. In all nine male equations the coefficient associated with years of school completed is positive and statistically significant;[18] on average, each year of schooling increases income nearly 5%, net of other factors. For women the return to education is much less consistent. In six of the nine countries, the return to education for women is less than that for men, and, indeed, in five of these countries—the United States constituting the only exception—the educational coefficient for women is not significantly different from zero. In Germany and Norway, in contrast, the education coefficients for females are substantially larger than those for males in any country and imply a 9% increase in income for each additional year of schooling. The variability in the return to education for men and women within each country calls seriously into question a simple differential investment explanation of the earnings gap.

Experience.—We expect a positive relationship between experience and earnings. Given what we know from U.S. data, however, we allow for the possibility that the relationship is curvilinear. Indeed, it is. In all 18 equations the slope of the squared term is negative, and in 16 of the 18 equations it is significantly so. Surprisingly, however, the linear term for experience behaves in a much less consistent way. Although most of the slopes are positive, only two of those for males and five for females are statistically significant. Thus, for males the picture is generally one of income increasing until the middle of the career and then falling as far as it has risen. For females, in contrast, in about half the countries the linear term for the experience variable is positive, although there is substantial curvature in the earnings trajectory for females as well. The

[18] In the following discussion, we refer to regression coefficients as statistically significant if they exceed twice their standard error. The exception is the set of dummy variables representing occupational categories, which we regard as significant if they significantly increase R^2 ($P < .05$).

TABLE 2

Occupational Distribution of Employed Persons, by Sex, for Nine Industrialized Countries (%)

Occupational Category*	Austria		Denmark		Finland		Federal Republic of Germany		Israel	
	F	M	F	M	F	M	F	M	F	M
1. High-prestige professional and technical	1.4	5.1	3.2	5.4	3.1	3.5	3.5	7.5	7.9	6.7
2. Administrative and managerial	.5	3.3	.0	3.8	1.2	2.4	.0	2.7	1.0	4.2
3. High-prestige clerical and related	12.7	3.7	24.1	3.8	14.8	.8	31.7	18.4	25.7	8.2
4. High-prestige sales	4.3	8.4	2.4	4.7	3.1	3.3	4.3	5.8	2.1	6.7
5. Low-prestige professional and technical	6.0	4.5	14.5	6.6	10.2	7.1	11.0	6.0	21.4	7.5
6. High-prestige agricultural	19.2	15.0	17.3	22.1	24.7	22.8	1.9	4.6	2.0	3.6
7. High-prestige production and related	3.5	16.7	3.6	15.3	3.1	10.6	5.3	23.3	5.0	19.6
8. High-prestige service	4.6	3.9	4.4	1.4	4.3	1.9	5.8	5.0	6.4	2.9
9. Medium-prestige production and related	7.3	19.8	3.6	22.4	6.2	28.3	4.6	18.2	4.9	18.5
10. Low-prestige clerical and related	10.6	7.6	2.8	2.6	4.6	2.7	9.1	4.5	3.0	4.0
11. Low-prestige sales	7.6	.8	6.4	1.9	6.5	.5	12.1	1.2	6.3	2.5
12. Low-prestige agricultural	1.9	3.3	.4	1.6	1.9	5.7	.7	.4	.7	2.3
13. Low-prestige service	12.5	.6	14.9	.5	13.3	1.4	5.1	.8	10.2	4.4
14. Low-prestige production and related	7.9	7.4	2.4	8.0	3.1	9.0	5.0	1.6	3.5	9.1
Total	100.0	100.1	100.0	100.1	100.1	100.0	100.1	100.0	100.1	100.2
N	342	541	281	393	319	373	386	642	4,407	8,791
Index of dissimilarity	37.6		50.4		41.2		42.6		45.8	

* Standard International Classification (Treiman 1977, chap. 9).

628

TABLE 2 (*Continued*)

OCCUPATIONAL CATEGORY*	NETHERLANDS		NORWAY		SWEDEN		United States	
	F	M	F	M	F	M	F	M
1. High-prestige professional and technical	6.3	7.3	5.8	8.3	6.2	8.5	6.7	11.1
2. Administrative and managerial	.5	8.0	2.3	4.8	.4	5.2	3.0	10.4
3. High-prestige clerical and related	27.3	11.7	18.1	6.0	26.3	2.4	22.9	2.4
4. High-prestige sales	3.9	10.0	1.2	5.0	1.1	4.7	1.5	4.4
5. Low-prestige professional and technical	18.0	9.0	13.5	12.1	17.2	10.7	15.6	6.5
6. High-prestige agricultural	.5	4.6	16.4	14.9	1.1	7.6	.1	3.4
7. High-prestige production and related	.5	12.9	.0	10.8	2.2	17.3	4.5	18.4
8. High-prestige service	3.4	4.4	2.9	3.5	5.5	3.8	8.8	5.0
9. Medium-prestige production and related	4.4	20.2	4.1	19.4	3.6	23.7	8.5	19.1
10. Low-prestige clerical and related	5.9	3.9	4.1	2.5	5.5	2.8	10.3	4.7
11. Low-prestige sales	11.2	1.5	12.9	1.5	10.6	2.1	5.3	5.1
12. Low-prestige agricultural	1.5	2.7	.6	2.0	.0	3.3	.4	1.7
13. Low-prestige service	15.1	1.2	15.2	1.3	18.2	1.4	10.4	2.9
14. Low-prestige production and related	1.5	2.7	2.9	7.8	2.2	6.4	2.0	5.0
Total	100.0	100.1	100.0	99.9	100.1	99.9	100.0	100.1
N	160	456	155	413	299	397	1,298	1,939
Index of dissimilarity	50.2		41.8		60.0		46.8	

* Standard International Classification (Treiman 1977, chap. 9).

TABLE 3

PERCENTAGE FEMALE IN EACH OCCUPATION GROUP EXPRESSED AS A DEVIATION FROM PERCENTAGE FEMALE IN THE TOTAL LABOR FORCE, FOR NINE INDUSTRIALIZED COUNTRIES

Occupational Category*	Austria	Denmark	Finland	Federal Republic of Germany	Israel	Netherlands	Norway	Sweden	United States	Nine-country Average†
1. High-prestige professional and technical	−24	−12	−3	−16	**+4**	−3	−6	−8	−11	−9
2. Administrative and managerial	−29	[−42]	[−16]	[−38]	−23	−24	−12	−38	−24	−25
3. High-prestige clerical and related	+30	+40	+48	+13	+28	+19	+26	+46	+46	+33
4. High-prestige sales	−14	−15	−1	−7	−20	−14	−19	−28	−21	−15
5. Low-prestige professional and technical	+7	+19	+9	+15	+26	+15	+2	+12	+22	+14
6. High-prestige agricultural	**+6**	−6	**+2**	+18	−11	−22	**+2**	−33	−39	−13
7. High-prestige production and related	−27	−27	−26	−25	−22	−25	−27	−34	−26	−27
8. High-prestige service	+4	[+27]	+20	+3	+19	**−4**	[−4]	+9	+14	+9
9. Medium-prestige production and related	−20	−31	−30	−24	−22	−19	−20	−32	−17	−24
10. Low-prestige clerical and related	+8	[+2]	+13	+17	**−6**	+9	[+11]	+16	+19	+11
11. Low-prestige sales	+47	+29	+45	+48	+22	+47	+49	+36	+1	+36
12. Low-prestige agricultural	−12	. . .	−24	. . .	−21	[−10]	. . .	[−43]	−28	−21
13. Low-prestige service	+54	+54	+43	+42	+21	+55	+55	+48	+30	+45
14. Low-prestige production and related	**+2**	−24	−23	**+28**	−17	[−10]	−15	−22	−19	−11
Total percentage female	39	42	46	38	33	26	27	43	40	37

NOTE.—Figures of sign opposite pattern for row are boldface. Percentages based on fewer than 10 cases not shown; percentages based on 10–19 cases shown in brackets.

* Standard International Classification (Treiman 1977, chap. 9).

† Bracketed percentages not included in average.

630

TABLE 4

MEAN AND STANDARD DEVIATION OF INCOME BY SEX, FULL-TIME WORKERS AGED 20-64, FOR NINE INDUSTRIALIZED COUNTRIES

COUNTRY	MEAN		FEMALE INCOME AS A PERCENTAGE OF MALE	RATIO OF MALE TO FEMALE INCOME	STANDARD DEVIATION		N CASES	
	Female	Male			Female	Male	Female	Male
Austria	4,586	6,932	66	1.51	2,337	3,368	216	475
Denmark	31,524	55,503	57	1.76	24,693	53,198	132	331
Finland	10,665	15,796	68	1.48	6,914	10,362	184	303
Federal Republic of Germany	1,674	2,276	74	1.36	711	1,019	169	520
Israel	11,335	17,388	65	1.53	6,072	8,391	1,147	3,175
Netherlands	1,028	1,569	66	1.53	605	927	91	386
Norway	25,047	39,818	63	1.59	11,097	22,249	78	377
Sweden	25,125	36,292	69	1.44	16,338	15,750	142	344
United States	7,333	14,243	51	1.94	4,880	9,312	810	1,515

NOTE.—Income given in local currency: Austria, Netherlands (net monthly earnings); Denmark, Finland, Norway, Sweden (gross annual income); Federal Republic of Germany (gross monthly earnings); Israel, United States (gross annual earnings).

TABLE 5

COEFFICIENTS OF A MODEL OF INCOME ATTAINMENT FOR FULL-TIME WORKERS AGED 20–64, BY SEX, FOR NINE INDUSTRIALIZED COUNTRIES
(Dependent Variable is Natural Log of Income)

	AUSTRIA		DENMARK		FINLAND		FEDERAL REPUBLIC OF GERMANY		ISRAEL	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
Metric coefficients										
Years of schooling	.0161	.0418	.0420	.0639	.0308	.0558	.0870	.0211	.0474	.0231
Years of potential experience	−.000245	.000238	.00154	.00431	.00669	.00271	.00886	.00275	.0154	.00110
Experience squared*	−.000694	−.000183	−.000868	−.000804	−.00103	−.000313	−.000646	−.000298	−.000618	−.000592
Never married	−.0130	−.227	.203	−.0939	−.0327	−.363	−.102	−.292	−.0899	−.247
Prestige of current occupation	.0000210	.00311	−.0292	.0137	−.0256	.00308	.0150	.00801	.0123	.00791
Administrative occupation†	.000748	.428	0.0	.0526	−.378	.0852	0.0	.211	.0876	−.0137
Clerical occupation	−.407	−.103	−.115	−.0936	−1.05	−.191	.214	−.188	−.0598	−.176
Sales occupation	−.350	.0354	−.758	−.186	−1.53	−.171	.177	.174	−.118	−.246
Service occupation	−.731	−.257	−1.23	−.0432	−2.20	−.380	.555	−.0890	−.144	−.229
Agricultural occupation	−1.04	−.563	−1.31	−.119	−1.99	−.705	.157	−.111	−.0430	−.361
Production occupation	−.762	−.130	−.945	−.104	−1.68	−.330	.171	−.166	−.200	−.250
Constant	8.89	8.42	11.7	9.74	11.2	9.36	5.53	7.16	8.11	9.44
Standard error of estimate	.440	.412	.594	.516	.447	.490	.387	.365	.575	.412
Standardized coefficients										
Years of schooling	.047	.168‡	.121	.225‡	.102	.188‡	.401‡	.113‡	.253‡	.185‡
Years of potential experience	−.005	.006	.029	.089	.116‡	.055	.253‡	.069	.308‡	.030
Experience squared*	−.179‡	−.152‡	−.190‡	−.188‡	−.206‡	−.077	−.222‡	−.092‡	−.148‡	−.214‡
Never married	−.010	−.166‡	.105	−.051	−.017	−.250‡	−.101	−.218‡	−.062	−.168‡
Prestige of current occupation	.000	.073	−.389‡	.263‡	−.403‡	.060	.349‡	.225‡	.221‡	.207‡
Occupation categories§	.588‡	.427‡	.773‡	.087	.975‡	.383‡	.252‡	.257‡	.101	.193‡
R²	.384	.379	.395	.301	.648	.387	.399	.313	.310	.364
Adj. R²	.350	.364	.337	.275	.622	.363	.357	.298	.304	.362

NOTE.—Means, SDs, and correlations given in Appendix B.
* Adjusted to be orthogonal to experience; see text for details.
† "Professional and technical" is the omitted category.
‡ Coefficient is at least twice its standard error.
§ Reiss's (1922) short coefficient. Significance measured by F-ratio for increment to R²

TABLE 5 (Continued)

	NETHERLANDS		NORWAY		SWEDEN		UNITED STATES	
	Female	Male	Female	Male	Female	Male	Female	Male
	Metric coefficients							
Years of schooling.	.0100	.0423	.0899	.0540	−.00164	.0533	.0525	.060
Years of potential experience.	.00831	.00588	−.00194	.00346	.00636	.00229	.0137	.0139
Experience squared*.	−.000552	−.000730	−.000954	−.000312	−.000188	−.000440	−.000393	−.00103
Never married.	−.0152	−.271	−.157	−.269	.160	−.0770	−.0343	−.300
Prestige of current occupation	.0224	.00816	−.0288	−.00105	.0197	.000852	.0191	.0119
Administrative occupation† . . .	−.359	.199	.280	.0623	−.644	.137	−.154	.316
Clerical occupation	.136	−.0240	−.356	−.0714	−.132	−.115	−.128	.0778
Sales occupation	.115	−.0391	−.782	−.127	−.265	−.107	−.171	.240
Service occupation	.289	−.0842	−1.29	.0300	−.689	−.0595	−.172	.0285
Agricultural occupation	.00408	−.192	−1.05	−.438	−2.41	−.418	−.367	.163
Production occupation	.196	−.157	−1.35	−.172	−.0490	−.151	.0214	.196
Constant	5.66	6.56	11.3	10.2	9.18	10.1	7.07	7.85
Standard error of estimate. . . .	.369	.323	.323	.389	.577	.298	.690	.601
	Standardized coefficients							
Years of schooling	.054	.265‡	.398‡	.330‡	−.006	.386‡	.176‡	.271‡
Years of potential experience. .	.237‡	.166‡	−.049	.097	.102	.076	.234‡	.253‡
Experience squared*.	−.184‡	−.264‡	−.293‡	−.107‡	−.037	−.178‡	−.081‡	−.228‡
Never married.	−.018	−.217‡	−.121	−.175‡	.078	−.068	−.016	−.141‡
Prestige of current occupation	.631‡	.230‡	−.616‡	−.029	.333‡	.029	.310‡	.214‡
Occupation categories§.	.244	.224‡	.932‡	.327‡	.482‡	.337‡	.101	.141‡
R².	.333	.551	.704	.329	.476	.427	.229	.315
Adj. R².	.240	.538	.644	.306	.422	.407	.218	.310

NOTE.—Means, SDs, and correlations given in Appendix B.
* Adjusted to be orthogonal to experience; see text for details.
† "Professional and technical" is the omitted category.
‡ Coefficient is at least twice its standard error.
§ Heise's (1972) sheaf coefficient. Significance measured by F-ratio for increment to R^2.

633

results for the United States, which show significant effects of both the linear and squared terms for both men and women, appear to be somewhat deviant relative to other industrialized countries.

Although the trajectory for males in the United States shows substantially greater curvature than that for females, this pattern does not hold consistently across countries; in some countries the coefficient of the squared term is larger for females. While it might be surmised that the curvilinear relation between experience and earnings reflects cohort variations in income, each cohort starting and ending with higher income than the previous cohort, this possibility is inconsistent with the substantial variability in the relationship between the male and female curves in different countries—unless rather complicated and specific historical circumstances can be shown to account for these variations. Exploration of this possibility must be left for future research.

Marital status.—As expected, married men tend to earn more than single men. In all nine countries the coefficient of the never-married dummy variable is negative, and in seven it is statistically significant. In these seven countries the cost for men of never having married is substantial. Whether because of the benefits or the responsibilities of marriage, married men on the average earn about 20% more than single men, net of other factors. This differential cannot be accounted for by their greater experience or generally higher-status occupations because these factors are included in the prediction equation. However, employers may prefer to hire and promote married men because they view them as more stable and reliable than single men.

For women, of course, we expect just the opposite, at least if the dual career is an important deterrent to income. Married women should earn less than single women, net of other factors. Hence, a good test of whether the dual career faced by married women is an important determinant of the male-female income gap is whether the coefficient of the marital status variable is significant for women. The evidence, however, is conclusively negative. In none of the nine countries is the coefficient of marital status significant for females, and in seven of the nine its sign is opposite that expected under the hypothesis. Interestingly, this result directly contradicts the findings of Mincer and Polachek (1978, p. 120), who boldly assert that "*all* studies, whether of domestic or international data, find that marital status and family characteristics affect female wages."

In interpreting our result, however, we must keep in mind the fact that the sample is restricted to full-time workers. As evidenced by their full-time participation in the labor force, these married women, whether by choice or necessity, have worked out an accommodation between the demands of work and family. Once the necessary arrangements to work full-time have been made, these women are evidently not penalized by their

marital status. Although the results described above suggest that the dual career has no effect on the earnings of married women who work full-time, it may be that it affects earnings indirectly by influencing which women participate in the labor force and the kind of work they do. These issues are investigated further in Roos (1981a).

Occupation.—Occupational status as we have measured it here has a somewhat complex relation to income, although for all countries and both sexes there is a significant net association of some sort. The nine countries exhibit several different patterns of relationship between occupation and income, and the country differences are not consistent for males and females. Among men in Denmark and women in Israel, the Netherlands, and the United States, the occupational categories have no effect on income net of prestige, but the net effect of prestige is substantial, averaging a 1.7% increase in income for each prestige point. For example, net of education, experience, and marital status, a mathematician would be expected to earn about 20% more than a laboratory technician, also a professional worker, on the basis of their 11-point difference in prestige. In contrast, among men in Austria, Finland, Norway, and Sweden and among women in Austria, prestige has no effect on income within the major occupational groups, but differences between the groups in average income net of other factors are substantial. Among men in Germany, Israel, the Netherlands, and the United States, and among women in Germany and Sweden, both prestige and occupational group membership influence income. This is true also of women in Denmark, Finland, and Norway, except that in those countries the net effect of prestige on income is negative. This last result can be interpreted as reflecting a situation in which prestige and income are regarded as alternative forms of occupational remuneration, so that when one is relatively high the other will be relatively low. Insofar as the major occupational groups strongly differentiate occupations with respect to their qualifications and requirements, the groups should be strongly differentiated with respect to prestige as well. In this circumstance we would expect the net effect of prestige on income to be negative. This is because the effect of prestige as an indicator of occupational qualifications and requirements will be largely absorbed by the dummy variables for occupational groups, leaving only the effect of prestige as an alternative form of remuneration.[19] The trade-off between prestige and income is hardly unfamiliar, although it is most commonly observed with respect to different kinds of jobs within specific occupations (e.g., the higher prestige and lower salaries of social scientists employed

[19] As is consistent with this claim, the proportion of variance in prestige explained by the occupational categories (the squared correlation ratio) is higher for women from Denmark and Norway than for any other groups (.71 in both cases) and is nearly as high for Finnish women (.65).

in universities compared with social scientists employed in nonacademic settings).

In general, the variables we have measured account for a modest portion of the variance in income among workers of a given sex in each country. For three of the 18 equations, however, the proportion of variance explained is much larger than the average of about a third. Among Finnish women, but not Finnish men, average income differences between occupational categories are extremely large, and professional workers in particular earn much more than members of other occupational groups. As a consequence, nearly two-thirds of the variance in income among Finnish women can be accounted for, almost all of it due to differences in occupational status and experience. The situation is similar for Norwegian women, except that the income differences among occupational groups are not as dramatic and schooling contributes importantly to income, as it does not do in Finland. The equation for Dutch men, which accounts for over half the variation in income, is quite different from those for the Finnish and Norwegian women; in the case of Dutch men all of the variables in the model contribute in the expected way to the explanation of income, and no single variable dominates the model. Clearly, much more work needs to be done to clarify the nature of national differences in the determinants of income, including attention in particular to institutional features of labor markets that are beyond the scope of the present analysis.

Disaggregating Gender Differences in Income

We turn now to our final task—disaggregating the observed difference in the average incomes of men and women workers into components representing the effect of differences in the rate of return to each determinant of income and components representing the effect of gender differences in the average levels of these determinants. We make use of a multiplicative decomposition (suggested by Michael Sobel, private communication [1981]), which we derive as follows. Suppose our regression model is:

$$\widehat{\ln Y} = a + \Sigma b_i X_i ,$$ (2)

and we estimate it separately for males and females. Then for the two sex groups it follows that

$$\overline{\ln Y}_m = a_m + \Sigma b_{im} \overline{X}_{im} ,$$ (3)

and

$$\overline{\ln Y}_f = a_f + \Sigma b_{if}\overline{X}_{if} .$$ (4)

Now, the difference between these two equations (eq. [3] − eq. [4]) can be expressed as

$$
\begin{aligned}
\overline{\ln Y}_m - \overline{\ln Y}_f &= (a_m - a_f) + \Sigma(b_{im}\overline{X}_{im} - b_{if}\overline{X}_{if}) \\
&= (a_m - a_f) + \Sigma\overline{X}_{im}(b_{im} - b_{if}) + \Sigma b_{im}(\overline{X}_{im} - \overline{X}_{if}) \quad (5) \\
&\quad + \Sigma[(\overline{X}_{if} - \overline{X}_{im})(b_{im} - b_{if})] .
\end{aligned}
$$

Or, equivalently, as

$$
\begin{aligned}
\exp(\overline{\ln Y}_m - \overline{\ln Y}_f) &= \exp(a_m - a_f) \cdot \Pi\exp[\overline{X}_{im}(b_{im} - b_{if})] \cdot \\
&\quad \Pi\exp[b_{im}(\overline{X}_{im} - \overline{X}_{if})] \cdot \qquad (6a) \\
&\quad \exp[\Sigma (\overline{X}_{if} - \overline{X}_{im})(b_{im} - b_{if})] \\
&= \exp(a_m - a_f) \cdot \Pi\exp[\overline{X}_{if}(b_{im} - b_{if})] \cdot \\
&\quad \Pi\exp[b_{if}(\overline{X}_{im} - \overline{X}_{if})] \cdot \qquad (6b) \\
&\quad \exp[\Sigma(\overline{X}_{im} - \overline{X}_{if})(b_{im} - b_{if})] .
\end{aligned}
$$

Furthermore, the left side of equation (6) can be shown to equal the ratio of the geometric means of the dependent variable. Thus the difference in the (geometric) mean income of men and women[20] can be decomposed into multiplicative components representing, in order, a part due to differences in the intercepts, a set of coefficients indicating gender differences in the rate of return to each independent variable, a set of coefficients indicating gender differences in average levels of the independent variables (composition), and a coefficient indicating the effect of interaction between differences in rates of return and differences in composition (see Winsborough and Dickinson [1971] for an analogous decomposition when the dependent variable is in linear form).

The coefficients in the decomposition tend to center on unity, as table 6 illustrates. The rate coefficient for each variable can be interpreted as indicating the net contribution (to the ratio of geometric means) of differences in the rate of return to that variable for men and women. These contributions were evaluated at two points, the male mean and the female mean, to produce two estimates for each coefficient (see eqq. [6a], [6b]).

[20] Although the arithmetic mean is more familiar, the geometric mean (which has the arithmetic mean as an upper bound) is an equally appropriate measure of the central tendency of a variable.

TABLE 6

DECOMPOSITION OF RATIO OF GEOMETRIC MEAN INCOME FOR MALE AND FEMALE FULL-TIME WORKERS AGED 20–64 IN EACH OF NINE INDUSTRIALIZED COUNTRIES

	AUSTRIA M	AUSTRIA F	DENMARK M	DENMARK F	FINLAND M	FINLAND F	FEDERAL REPUBLIC OF GERMANY M	GERMANY F	ISRAEL M	ISRAEL F	NETHERLANDS M	NETHERLANDS F	NORWAY M	NORWAY F	SWEDEN M	SWEDEN F	UNITED STATES M	UNITED STATES F
Ratio of geometric means...	1.53		1.77		1.55		1.37		1.63		1.49		1.61		1.66		1.99	
Effect of sex differences in rates of return to independent variables:*	1.58	1.40	2.00	1.68	1.65	1.56	1.17	1.23	1.61	1.66	1.40	1.31	1.86	1.55	1.93	1.77	1.90	1.78
Intercept	.63		.14		.15		5.10		3.77		2.46		.32		2.41		2.18	
Years of schooling	1.29	1.27	1.19	1.20	1.22	1.22	.48	.49	.79	.77	1.39	1.40	.72	.73	1.62	1.65	1.11	1.11
Years of experience	1.01	1.01	1.08	1.07	.91	.89	.86	.88	.70	.78	.95	.97	1.16	1.15	.89	.91	1.00	1.00
Experience squared	1.10	1.07	1.01	1.01	1.12	1.12	1.04	1.07	1.01	1.01	.97	.97	1.12	1.12	.96	.96	.89	.89
Never married	.96	.95	.96	.95	.93	.94	.98	.94	.98	.95	.96	.87	.99	.98	.97	.96	.96	.96
Prestige of current occupation	1.13	1.12	5.54	5.48	3.06	3.01	.73	.74	.84	.83	.54	.54	3.26	3.16	.46	.45	.74	.73
Occupational group	1.61	1.52	2.07	1.78	3.11	3.03	.75	.75	.93	.92	.86	.86	1.93	1.66	1.30	1.18	1.24	1.18
Effect of sex differences in means on independent variables:	1.09	.97	1.01	.86	.97	.94	1.12	1.17	.99	1.01	1.12	1.08	1.04	.87	.93	.87	1.11	1.04
Years of schooling	1.03	1.01	.98	.99	.99	.99	1.01	1.03	.98	.95	.99	1.00	1.01	1.02	.98	1.00	1.01	1.01
Years of experience	1.00	1.00	1.01	1.01	.99	.98	1.01	1.04	1.01	1.11	1.05	1.08	1.01	1.00	1.01	1.02	1.01	1.01
Experience squared	.99	.97	1.03	1.03	1.00	1.00	1.02	1.05	1.00	1.01	1.00	1.00	1.00	1.01	.99	1.00	1.01	1.00
Never married	1.01	1.00	1.00	.99	.98	1.00	1.07	1.02	1.05	1.02	1.10	1.01	1.03	1.02	1.00	1.00	1.01	1.00
Prestige of current occupation	1.01	1.00	1.00	.99	1.00	.99	1.02	1.03	.99	.98	1.01	1.02	1.00	.97	1.00	.98	1.00	1.00
Occupational group	1.05	.99	.99	.85	1.01	.98	.99	.99	.96	.96	.97	.97	.99	.85	.95	.87	1.07	1.02
Interaction	.89	1.12	.84	1.19	.97	1.03	1.04	.96	1.03	.97	.96	1.05	.83	1.21	.92	1.09	.94	1.06

NOTE.—M indicates that males were used as the standard population (eq. [5a]), F indicates that females were used as the standard population (eq. [5b]).

* Includes intercept.

As can be seen in table 6, the two sets of estimates (labeled M and F) are generally very similar. The composition coefficient for each variable can be interpreted as indicating the net contribution of differences in the mean levels of the variable for males and females. Again, the effect of these differences was evaluated twice, once using the male rate of return and once using the female rate, and again the estimates are generally quite similar. The interaction component gives the effect of sex differences in rates of return and in composition considered simultaneously. It can be thought of as an inflator or deflator for the effect of differences in rates and in composition considered singly. Finally, the intercept coefficient gives the ratio of geometric mean income that would be expected if the value of each of the independent variables were zero. Like intercept terms in regression equations, it can be interpreted only if a true zero point is established, as we have not been able to do here. In the present case the intercept term also includes the effect of the omitted occupation category, professional and technical workers.

What does table 6 tell us about the determinants of income differences between male and female workers in the nine countries? First, we confirm that the ratios of geometric means are very close to the ratios of arithmetic means shown in table 4, so we are decomposing familiar quantities and thus should have no difficulties of interpretation. Note that the ratios range from 1.37 for Germany, indicating that on the average men earn about a third again as much as women, to 1.99 for the United States, indicating that on the average men earn about twice as much as women.

Inspecting the italicized summary rows of the table, we see that the greater income of men in each country is due mainly to their higher rates of return on the variables included in the model rather than to gender differences in the mean levels of these variables. The average rate effect (for the 18 columns) is 1.61, which is virtually identical to the average ratio of the geometric means for the nine countries (1.62). In contrast—and implied by the size of the rate effect—the average composition effect is 1.01, which indicates that gender differences in education, experience, marital status, and occupational position taken together do not contribute to the income gap at all.

Germany and the Netherlands are, however, exceptions to this pattern. In Germany the effect of composition is nearly as large as the rate effect, and in the Netherlands there is also a smaller than average rate effect and a larger than average composition effect. This appears to be due largely to the fact that the female labor force in these two countries is very young, since in both countries women tend to work only until they marry and then to leave the labor force permanently (see Appendix B).

The disaggregated coefficients show clearly that the greatest payoffs for men are associated with education and occupation. Consider education

first. In six of the nine countries the returns to schooling are substantially greater for men, averaging more than 30% higher. In Israel and Norway, however, and most notably in Germany, the returns to education are substantially smaller for men than for women. Recall that the return to schooling is higher for German and Norwegian women than for any other group, about 9% per year of schooling. Note also from table 5 that the return to education is lower for German and Israeli men than for men in any other country. What accounts for these patterns is unclear and warrants further investigation.

Gender differences in return to occupational position have already been anticipated in our discussion of table 5. Further elaboration is somewhat cumbersome since the coefficient associated with occupational group cannot be interpreted independently of the intercept coefficient because both coefficients are affected by the choice of omitted category in the regression estimation. Here it suffices to note that the product of the intercept term and the two occupation terms exceeds unity in all 18 sets of coefficients, averaging 1.78. Thus, in all countries men enjoy a greater return than women on either the prestige of their occupation or their occupational group membership, or both, and on average the difference in the rate of return is very substantial. In contrast, sex differences in occupational composition have no effect on the income gap in most countries. This fact undercuts a major hypothesis—that women are paid poorly because they are segregated into low-paying occupational sectors. While women may be segregated into low-paying *jobs*, differences in the distribution of men and women over the *major occupation groups* have virtually no effect on income. Sex differences in income are on average as large within each of the major occupation groups as they are for the labor force as a whole.

On the whole, neither experience nor marital status contributes much to the gender gap in income. Sex differences in the rate of return to experience vary from country to country in a not very systematic way and are never very large. Sex differences in the return to marital status are consistent but very modest: men benefit more than women because, as we saw in table 5, men experience a positive return to marriage while marital status has no effect on income for working women (the coefficients are less than unity because of the way the variable was coded). With the exception of Germany and the Netherlands, noted above, the effect of compositional differences in these variables is negligible. The lack of compositional differences for the experience variable may, of course, be due in part to the inadequacy of our measures.

DISCUSSION

This analysis represents a first attempt to investigate within a comparative framework why women workers in other industrial societies, as in the

United States, earn substantially less than their male counterparts. Four major hypotheses regarding gender differences in earnings, drawn from the American literature, were considered: (1) women have less education and labor force experience than men; (2) women's family responsibilities affect the time and energy they can devote to advancing within their employment; (3) women do different work from men, and the work they do pays less; and (4) women are subject to discriminatory market mechanisms that affect their earnings relative to men's. In this concluding section, we provide a summary of our major findings and assess the extent to which the data support the various hypothesized explanations for the gender differential in earnings. Only partial tests of these hypotheses could be made because of limitations in the kinds of variables available for analysis in the various countries. The analysis, does, however, address each of the explanations, with varying levels of success.

Our data show quite convincingly that men and women in industrial society work in very different kinds of jobs. Although the sex-segregated nature of the labor market is well-documented in the United States, hitherto there has been little systematic evidence that the same phenomenon occurs in other countries as well (see Gaskin 1979; Roos 1981a). We have shown not only that the general U.S. pattern of substantial occupational segregation by sex is characteristic of other industrial societies as well, but also that the degree of similarity across countries is striking. In all nine countries women are substantially overrepresented in high-prestige clerical occupations and in low-prestige sales and service employment; they are substantially underrepresented in administrative occupations and in high- and medium-prestige production occupations. Given these large and systematic gender differences in occupational distribution, any investigation of income attainment would be remiss if it did not attempt to incorporate measures of occupational segregation into the earnings equations.

The results of this study confirm unambiguously for the remaining countries what we already know about the United States—that women earn substantially less than men. Moreover, adjusting for gender differences in composition with respect to several variables thought to affect earnings failed in general to reduce the gap substantially. The adjusted earnings of women—what women would earn if on average they had the same characteristics as men—never exceeded a 17% increment over their actual earnings. The figure of 17% is the high estimate for the Federal Republic of Germany, which is the country for which the actual earnings of women came closest to those of men (see the summary row in table 6, "Effect of sex differences in means on independent variables"). In our data for Germany, male income exceeds female income by 37% (see the top row of table 6), so even for Germany the adjustment accounts for less than half of the earnings gap. At the other extreme, in the four Nordic

countries, adjusting for sex differences in composition would actually slightly reduce the expected earnings of women and hence increase the earnings gap. In the United States, where—in these data—men earn nearly twice as much as women on the average, adjusting for sex differences in occupational composition (the only variable of importance) would increase the earnings of women less than 10%.

These results require us to reject the first three hypotheses reviewed above—the human capital, occupational segregation, and dual burden hypotheses. The first two hypotheses each posit differences in the composition of the male and female labor force and similarities in rates of return to "investments." The human capital hypothesis suggests that men earn more than women because they have more education and experience but assumes that the rate of return to these characteristics will be similar for men and women, net of other identified factors. As we have seen, the effect of differences in average education and experience is very small (although, to be sure, the lack of difference in measured experience no doubt reflects the inadequacy of our measure) and the effect of differences in returns to education (although not to experience) is quite large. The occupational segregation hyopthesis posits that women are paid less than men because they are concentrated in low-paying sectors of the occupational structure but they enjoy equal returns to their occupational status. As we have seen, however, there is almost no effect of sex differences in average occupational prestige or in distribution over major occupation groups, but there are substantial, and complicated, differences in the rate of return to these characteristics for men and women. To be sure, this may be due in part to the crudeness of our classification. A more detailed occupational classification might show part of the gender difference in income to be due to the concentration of women in low-paying occupations, as has been suggested for the United States (Roos 1981b).

The dual career hypothesis must be rejected on somewhat different grounds. Our test of it requires that ever-married women exhibit a reduction in income relative to never-married women. This proves not to be the case: marriage has no impact on the earnings of women in the full-time labor force. It does, however, substantially benefit men. We regard these results as requiring more careful specification of "dual career" explanations for the earnings gap between men and women. Insofar as the dual demands of work and family make it difficult for women with families to maximize their earnings potential, we would have expected married women to earn less than single women. The fact that they do not suggests that, at least for those women who have been able to combine family life with full-time employment, family obligations do not inhibit earning power. Of course, married women are much less likely to work full-time; hence, full-time workers are a more highly selected subset of

married women than of single women. The major impact of marriage on earnings is apparently indirect, through its impact on labor force participation.

If the human capital, dual burden, and occupational segregation hypotheses cannot account for any substantial fraction of the earnings gap between men and women workers, what can we conclude? The failure of our data to support these hypotheses may certainly be due in part to the inadequacy of our measures. In particular, our reliance on a major group occupational classification means that we have been unable to measure occupational segregation as we would have preferred—at a relatively disaggregated level. The measurement decision we took has the consequence of including in our residual category those sex differences in earnings attributable to job (or even occupational) segregation. The earnings gap that remains thus includes such factors as lack of equal pay for the same job (wage discrimination) and the differential concentration of men and women in different jobs and occupations within major occupation groups and levels of prestige (whether the allocation occurs because of discrimination or self-selection). A more disaggregated occupational classification would thus help to estimate the relative importance of these two explanations for the earnings gap. However, while our failure to support any of the three hypotheses may be due in part to the crudeness of our measures, a more likely explanation is that the observed income differences reflect the legacy of traditional patterns of disadvantage and discrimination faced by working women, patterns that were legitimated and reinforced by law in most industrialized countries until very recently and that continue to pervade labor market institutions. In addition, of course, they may reflect deeply ingrained cultural definitions of appropriate behavior for men and women, definitions that are transmitted from generation to generation by both parental and societal socialization processes. While both institutional arrangements and cultural definitions are changing in many countries, the process of change is a slow one. Hence, closer attention to both structural and cultural factors specific to each country may help us to understand better both the gross pattern of similarity in all industrial societies that we have observed and the departures from this general pattern that we have found puzzling. This is the obvious next step in research on this topic.

APPENDIX A

PERCENTAGE EMPLOYED FULL-TIME BY AGE AND SEX, FOR NINE INDUSTRIALIZED COUNTRIES

Age	Austria	Denmark	Finland	Federal Republic of Germany	Israel	Netherlands	Norway	Sweden	United States
Females:									
< 20........	26.1	23.7	21.9	53.1	5.3*	36.5	11.4	12.1	23.3
20–24.......	58.0	56.9	54.7	51.0	27.0†	72.0	23.5	33.3	40.2
25–29.......	33.8	42.2	55.6	31.4	23.9	31.6	21.0	33.3	37.9
30–34.......	28.0	35.6	58.5	20.0	17.5	18.0	20.3	33.3	36.9
35–44.......	42.5	33.7	78.0	18.2	15.6	13.3	21.7	31.7	35.8
45–54.......	42.6	35.3	65.2	14.9	19.4	15.3	27.3	35.5	36.5
55–64.......	21.5	23.6	49.0	13.6	10.1	8.0	20.3	12.9	30.5
65 + ...	2.7	...	...	2.0	2.3	3.0	...	...	3.1
Total......	31.5	34.1	55.0	18.7	16.0	25.0	21.4	28.0	30.3
Males:									
< 20........	41.9	31.9	25.8	50.3	11.7*	31.8	13.5	14.6	32.5
20–24.......	89.7	75.7	65.2	66.6	31.4†	67.9	57.8	53.3	57.6
25–29.......	96.6	86.0	89.3	74.0	72.0	91.8	89.3	85.7	75.3
30–34.......	98.8	92.1	88.7	87.5	76.2	91.5	96.1	90.9	80.8
35–44.......	98.6	100.0	92.3	97.5	79.4	93.7	94.7	92.3	81.1
45–54.......	96.4	88.9	78.5	94.9	81.6	85.5	92.1	93.1	80.1
55–64.......	62.5	82.8	56.2	66.4	73.5	70.6	70.9	77.5	60.1
65 + ...	17.2	...	...	5.3	17.8	5.8	...	...	8.9
Total......	79.5	83.0	70.9	72.9	54.4	68.5	77.5	77.8	60.8

* 14–17.
† 18–24.

APPENDIX B: MEANS, STANDARD DEVIATIONS, AND CORRELATIONS OF ALL VARIABLES FOR NINE INDUSTRIALIZED COUNTRIES

Austria

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Means Female	Means Male	Standard Deviation Female	Standard Deviation Male
1. Years of schooling		-.212	.080	.081	.404	.437	.152	.349	.092	-.191	-.387	.187	.404	.417	9.19	9.93	1.58	2.08
2. Years of potential experience[a]	-.203		.009	-.303	-.038	-.074	.079	-.231	.074	.067	.154	-.019	-.099	-.108	26.1	24.4	11.8	12.4
3. Experience squared[a]	-.128	-.002			.040	.104	-.001	.029	.050	.092	-.115	-.106	-.099	-.089	140	179	140	430
4. Never married	.046	-.394	.060	.155	-.090	.055	-.046	.122	-.026	.074	-.147	-.004	-.012	.043	.228	.172	.420	.370
5. Prestige of current occupation	.558	-.042	-.014	-.081		.480	.199	.193	-.008	-.410	.128	-.348	.315	.234	36.9	39.3	10.8	12.1
6. Professional occupation	.462	-.074	-.031	.005	.525		-.046	-.155	-.039	-.127	-.172	-.140	.204	.222	.083	.093	.276	.291
7. Administrative occupation	.246	-.052	-.014	-.057	.366	-.061		-.044	-.032	-.036	-.049	-.040	.116	.104	.007	.035	.085	.184
8. Clerical occupation	.023	.013	-.008	.030	.066	-.114	-.068		-.114	-.218	-.294	-.239	.150	.223	.209	.112	.408	.315
9. Sales occupation	.084	-.026	.142	-.057	.184	-.070	-.037	-.107	-.196	-.161	-.217	-.176	.229	.192	.126	.063	.333	.276
10. Service occupation	.037	.018	-.008	-.047	-.104	-.096	.042	-.078	-.046		-.241	-.196	-.153	-.104	.152	.046	.339	.209
11. Agricultural occupation	-.256	.199	.071	.027	-.095	-.151	-.090	-.167	-.142	-.103		-.244	-.227	-.428	.244	.181	.431	.386
12. Production occupation	-.217	-.131	-.101	.030	-.431	-.290	.173	-.320	-.271	-.197	-.424		-.177	-.101	.177	.449	.382	.494
13. Income in local currency	.463	.028	-.112	-.190	.435	.211	.430	-.012	.148	-.036	.298	-.131		.894	4586	6932	2337	3348
14. Log income	.401	-.028	-.178	-.179	.341	.177	.300	-.044	.133	-.031	-.412	.005	.905		8.30	8.73	.546	.517
N															216	473		

Denmark

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Means Female	Means Male	Standard Deviation Female	Standard Deviation Male
1. Years of schooling		-.387	.129	.268	.650	.443	[c]	.221	-.056	-.261	-.287	-.163	.339	.296	8.43	8.11	2.11	2.13
2. Years of potential experience[a]	-.199		.001	-.267	-.261	-.265	-	-.197	.178	-.003	.358	-.016	-.159	-.243	24.8	24.2	13.0	12.5
3. Experience squared[a]	-.053	-.002		.164	.097	.042	-	-.001	-.047	-.126	.074	.029	-.123	-.174	191	156	160	161
4. Never married	-.125	-.170	.150		.190	.179	-	.018	-.068	.002	-.168	.024	.088	.159	.172	.125	.378	.331
5. Prestige of current occupation	.649	-.057	.020	-.146		.585	-	.318	-.039	-.570	-.229	-.225	.264	.273	39.7	39.9	9.74	11.6
6. Professional occupation	.615	-.085	-.026	-.115	.556		-	-	-.148	-.214	-.281	-.168	.039	.102	-.210	.120	.409	.325
7. Administrative occupation	.267	-.013	-.057	-.078	.355	-.076		-	-	-	-	-	-	-	-	.041	-	.198
8. Clerical occupation	.005	-.012	.049	.054	-.003	-.098	-.055		.163	-.234	-.308	-.184	.178	.320	.242	.066	.430	.249
9. Sales occupation	-.000	-.043	-.008	.111	.070	-.100	-.056	-.072		-.119	-.157	-.094	.108	-.024	.076	.069	.266	.253
10. Service occupation	-.025	-.022	-.046	.000	-.119	-.053	-.030	-.038	-.039		-.226	-.135	-.150	-.113	.146	.020	.355	.141
11. Agricultural occupation	-.233	.284	.100	.052	-.004	-.211	-.118	-.152	-.135	-.063		-.177	-.253	-.456	.229	.247	.422	.432
12. Production occupation	-.239	-.194	-.053	-.022	-.509	-.325	-.182	-.235	-.240	-.127	-.505		-.011	-.054	.096	.438	.293	.497
13. Income in local currency	.332	.013	-.138	-.099	.394	.239	.168	-.056	-.040	-.031	.088	-.242		.779	3124	55503	24692	55196
14. Log income	.444	.028	-.208	-.175	.452	.335	.216	-.013	-.024	-.018	-.073	-.224	.853		10.1	10.7	.730	.606
N															132	331		

Finland

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Means Female	Means Male	Standard Deviation Female	Standard Deviation Male
1. Years of schooling		-.359	-.075	.051	.653	.591	.251	.112	-.074	-.177	-.277	-.147	.555	.427	8.08	7.86	2.41	2.07
2. Years of potential experience[a]	-.300		.053	-.142	-.097	-.123	-.037	-.104	-.076	.023	.267	-.090	-.052	-.078	28.4	25.0	12.6	12.5
3. Experience squared[a]	-.073	.017	.193	.130	-.090	-.044	-.009	.058	-.063	.082	-.059	-.038	-.214	-.208	157	159	146	150
4. Never married	-.078	-.295	.040		.056	.084	.024	.170	-.067	.031	-.163	-.038	.038	.073	.176	.232	.381	.423
5. Prestige of current occupation	.476	-.039	-.045	-.122		.564	.235	.192	-.034	-.511	-.093	-.260	.491	.341	38.4	38.9	11.4	11.9
6. Professional occupation	.377	-.070	.022	-.070	.503		-.049	-.184	-.119	-.152	-.267	-.157	.444	.373	.134	.126	.341	.332
7. Administrative occupation	.411	-.024	.001	-.095	.313	-.066		-.059	-.038	-.048	-.065	-.050	.139	.116	.013	.029	.123	.168
8. Clerical occupation	.022	-.044	.171	-.064	-.047	-.073	-.033		-.042	-.183	-.322	-.189	-.072	.184	.183	.036	.388	.185
9. Sales occupation	.089	-.107	.019	.064	.114	-.082	-.038	-.035	-.040	-.183	-.206	-.111	-.075	-.011	.084	.045	.277	.208
10. Service occupation	.085	-.057	.019	-.057	-.135	-.069	-.032	-.055	-.126	.117	-.263	-.154	-.210	-.214	.130	.032	.337	.177
11. Agricultural occupation	-.199	.228	.121	.121	.063	-.220	-.100	-.111	.040	.106		-.272	-.219	-.448	.317	.252	.466	.433
12. Production occupation	-.293	-.114	-.158	-.009	-.496	-.363	-.166	-.185	.209	-.176	-.558		-.137	-.040	.137	.481	.343	.500
13. Income in local currency	.442	.030	-.105	-.265	.419	.330	.333	.088	.031	-.028	-.330	-.123		.863	10645	13776	6934	10542
14. Log income	.386	.001	-.138	-.357	.288	.332	.245	.073	.056	.012	-.386	-.029	.854		9.06	9.49	.727	.614
N															164	391		

NOTE.—Males below the diagonal, females above.

[a] Adjusted to be orthogonal to experience; see text for details.

[b] Missing cases deleted pairwise; lowest number of cases reported.

[c] There are no female administrators in the sample.

645

APPENDIX B (*Continued*)

Germany (Fed. Rep.)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Means Female	Means Male	S.D. Female	S.D. Male
1. Years of schooling		-.329	-.057	.113	.427	.469	--c	.022	.101	.169	.130	.197	.419	.374	10.8	11.1	2.22	2.34
2. Years of potential experience	-.262		.015	-.450	-.151	-.107		-.189	.007	.144	.195	.139	.170	.143	20.2	24.2	11.8	10.9
3. Experience squared	-.021	.048		.266	-.144	-.172		.124	-.031	.131	.022	-.076	-.264	-.246	.184	.120	.166	.134
4. Never married	-.025	-.434	.391		-.041	-.008		.067	-.037	.010	-.106	-.004	-.255	-.237	.347	.119	.477	.324
5. Prestige of current occupation	.612	-.100	-.026	-.079		.434		.268	-.094	-.509	-.093	-.448	.482	.414	42.5	44.4	11.2	12.2
6. Professional occupation	.501	-.135	-.017	-.021	.507			-.385	-.164	-.112	-.080	-.184	.184	.161	.151	.134	.360	.341
7. Administrative occupation	.269	-.014	-.010	-.015	.283	-.067			-.357							.028		.165
8. Clerical occupation	-.009	-.010	-.027	-.023	.192	-.216	-.093			-.244	-.173	-.400	-.105	-.064	.458	-.232	.499	.422
9. Sales occupation	-.025	-.040	-.004	-.028	-.035	-.103	-.045	-.144		-.103	-.073	-.149	-.118	-.072	.131	.064	.338	.244
10. Service occupation	-.078	.014	.043	.107	-.152	-.099	-.043	-.138	-.046		-.050	-.115	-.032	-.024	.066	.059	.248	.236
11. Agricultural occupation	-.106	.057	.017	.038	-.048	-.088	-.038	-.123	-.059	-.056		-.062	-.013	-.042	.034	.047	.182	.213
12. Production occupation	-.359	.110	-.016	-.062	-.531	-.345	-.149	-.482	-.230	-.220	-.196		-.231	-.193	.159	.435	.346	.496
13. Income in local currency	.442	.035	-.133	-.233	.444	.307	.276	-.078	.259	-.066	-.034	-.308		.925	1674	2276	711	1019
14. Log income	.343	.079	-.138	-.297	.408	.270	.222	-.061	.179	-.043	-.045	-.251	.903		7.32	7.64	.442	.456
R²															169	520		

Israel

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Means Female	Means Male	S.D. Female	S.D. Male
1. Years of schooling		-.425	-.148	.015	.642	.449	.075	.157	.134	.396	.128	.230	.433	.339	10.6	9.56	3.67	4.11
2. Years of potential experience	-.419		.000	-.529	-.257	-.119	.021	-.229	.208	.267	.086	-.007	.123	.158	17.7	24.3	13.8	14.0
3. Experience squared	-.098	.003		.180	-.109	-.059	-.054	-.002	-.009	.067	-.043	.056	-.282	-.231	206	198	165	186
4. Never married	.021	-.435	.306		-.072	-.093	-.036	.128	-.098	-.077	-.058	.120	-.284	-.272	.353	.142	.478	.350
5. Prestige of current occupation	.577	-.156	-.082	-.076		.554	.186	.178	-.099	-.504	-.067	-.344	.483	.400	41.5	35.8	12.4	13.5
6. Professional occupation	.473	-.197	-.005	-.010	.464		-.059	-.421	-.163	-.222	-.013	-.232	.345	.285	.224	.137	.423	.344
7. Administrative occupation	.160	.047	-.085	-.084	.361	-.087		-.062	-.032	-.043	-.015	-.045	.133	.104	.011	.043	.104	.208
8. Clerical occupation	.139	.044	-.045	-.013	.052	-.151	-.083		-.224	-.307	-.105	-.320	-.000	.024	.348	.126	.482	.332
9. Sales occupation	.001	.073	.036	-.023	-.060	-.124	-.040	-.110		-.118	-.041	-.123	-.011	-.000	.080	.088	.271	.284
10. Service occupation	-.166	.128	.054	-.016	-.249	-.110	-.083	-.105	-.086		-.056	-.185	-.179	-.172	.139	.071	.346	.257
11. Agricultural occupation	-.142	.063	.025	.007	-.052	-.096	-.040	-.092	-.075	-.067		-.058		-.055	.039	.055	.136	.229
12. Production occupation	-.323	-.048	.006	.070	-.448	-.301	-.263	-.363	.297	-.264	-.231		-.282	-.229	.149	.477	.356	.500
13. Income in local currency	.433	-.057	-.282	-.159	.483	.276	.278	-.034	-.021	-.126	-.102	-.253		.848	1335	1788	6072	8191
14. Log income	.416	-.035	-.310	-.269	.431	.265	.216	.047	-.015	-.103	-.099	-.260	.902		9.16	9.65	.688	.513
R²															1147	3175		

Netherlands

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Means Female	Means Male	S.D. Female	S.D. Male
1. Years of schooling		-.273	.034	.077	.459	.363	.072	.013	.239	-.210	.136	.003	.192	.158	10.4	10.1	2.29	2.98
2. Years of potential experience	-.450		-.003	-.318	.054	-.025	.172	.140	.138	.074	.104	-.110	.205	.249	14.1	25.1	12.1	13.4
3. Experience squared	-.019	.007		.141	.097	.116	.042	-.108	.039	-.020	-.091	-.019	-.120	-.147	192	185	141	172
4. Never married	.122	-.422	.285		.032	.131	-.097	.032	-.207	.016	-.097	.059	-.083	-.094	.536	.176	.501	.381
5. Prestige of current occupation	.546	-.092	-.100	.102		.600	.170	.095	-.199	-.591	-.020	-.233	.365	.466	42.6	43.7	11.9	13.4
6. Professional occupation	.561	-.192	-.019	.027	.512		-.057	.461	-.278	-.261	-.057	-.143	.156	.192	.284	.173	.455	.379
7. Administrative occupation	.091	.079	-.153	.144	.471	-.142		-.065	-.039	-.037	-.008	-.020	.015	.042	.008	.088	.090	.284
8. Clerical occupation	.001	.008	.087	.077	-.071	-.194	-.132		-.316	-.297	-.065	-.163	-.004	.086	.344	.152	.478	.360
9. Sales occupation	-.062	.053	.034	-.002	-.043	-.169	-.113	-.156		-.179	-.039	-.092	-.004	-.108	.160	.120	.368	.325
10. Service occupation	-.110	.087	-.002	-.043	-.071	-.106	-.072	-.098	-.105		-.037	-.092	-.140	-.192	.144	.051	.353	.220
11. Agricultural occupation	-.201	.125	.017	-.051	-.159	-.130	-.088	-.120	-.105	-.044		-.020	-.108		.008	.075	.090	.263
12. Production occupation	-.299	-.046	-.021	-.096	-.494	-.330	-.224	-.303	-.266	-.146	-.204		.326	.126	.048	.341	.215	.475
13. Income in local currency	.396	.095	-.282	-.287	.546	.241	.398	-.034	.013	-.072	-.116	-.358		.858	1024	1549	605	927
14. Log income	.383	.113	-.310	-.378	.566	.289	.389	-.020	.011	-.056	-.140				6.83	7.23	.423	.475
R²															91	386		

APPENDIX B (Continued)

Norway

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Means Female	Means Male	Std. Dev. Female	Std. Dev. Male
1. Years of schooling																		
2. Years of potential experience																		
3. Experience squared																		
4. Never married																		
5. Prestige of current occupation																		
6. Professional occupation																		
7. Administrative occupation																		
8. Clerical occupation																		
9. Sales occupation																		
10. Service occupation																		
11. Agricultural occupation																		
12. Production occupation																		
13. Income in local currency																		
14. Log income																		

Sweden

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Means Female	Means Male	Std. Dev. Female	Std. Dev. Male
1. Years of schooling																		
2. Years of potential experience																		
3. Experience squared																		
4. Never married																		
5. Prestige of current occupation																		
6. Professional occupation																		
7. Administrative occupation																		
8. Clerical occupation																		
9. Sales occupation																		
10. Service occupation																		
11. Agricultural occupation																		
12. Production occupation																		
13. Income in local currency																		
14. Log income																		

United States

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Means Female	Means Male	Std. Dev. Female	Std. Dev. Male
1. Years of schooling																		
2. Years of potential experience																		
3. Experience squared																		
4. Never married																		
5. Prestige of current occupation																		
6. Professional occupation																		
7. Administrative occupation																		
8. Clerical occupation																		
9. Sales occupation																		
10. Service occupation																		
11. Agricultural occupation																		
12. Production occupation																		
13. Income in local currency																		
14. Log income																		

REFERENCES

Astin, Helen S., and Alan E. Bayer. 1975. "Sex Discrimination
in *Women and Achievement: Social and Motivational Analy*
Mednick, Sandra S. Tangri, and Lois W. Hoffman. Washingt
Blau, Francine D. 1977. *Equal Pay in the Office*. Lexington, M
Blau, Francine D., and Wallace E. Hendricks. 1979. "Occupat
Trends and Prospects." *Journal of Human Resources* 14:197–i
Blau, Francine D., and Carol L. Jusenius. 1976. "Economists'
gation in the Labor Market: An Appraisal." Pp. 181–99 in *H*
The Implications of Occupational Segregation, edited by Ma
Reagan. Chicago: University of Chicago Press.
Boyd, Monica, and Elizabeth Humphreys. 1979. "Sex Difference
Labour Markets." Paper presented at the Economic Council
Incomes, Winnipeg, Manitoba, May 9–12.
Bridges, William P., and Richard A. Berk. 1974. "Determinant
An Evaluation of Equal Pay for Equal Work." *Social Science*
Buckley, J. 1971. "Pay Differences between Men and Women ir
Labor Review 94:36–39.
Durand, John D. 1975. *The Labor Force in Economic Devel.*
International Census Data, 1946–1966. Princeton, N.J.: Princ
England, Paula. 1979. "Women and Occupational Prestige: A Casi
Signs 5:252–65.
————. 1982. "The Failure of Human Capital Theory to Expla
regation." *Journal of Human Resources* 17:358–70.
Federal Republic of Germany, Federal Institute for Population
ulation of the Federal Republic of Germany. Wiesbaden: Feder
Research.
Gaskin, Katharine A. 1979. "Occupational Differentiation by Si
parison." Ph.D. dissertation, University of Michigan.
Gross, Edward. 1968. "Plus Ça Change . . . ? The Sexual Stru
Time." *Social Problems* 16:198–208.
Haavio-Mannila, Elina, and Magdalena Sokolowska. 1978. "S
Pp. 183–216 in *Social Structure and Change: Finland and Pi*
spective, edited by Erik Allardt and Wlodzimierz Wesolowski.
Publishers.
Heise, David R. 1972. "Employing Nominal Variables, Induced
ables in Path Analyses." *Sociological Methods and Research* 1
International Labour Office. 1969. *International Standard Clai*
Rev. ed. Geneva: International Labour Office.
————. 1977. *Yearbook of Labour Statistics.* Geneva: Internatio
Israel Labor Mobility Survey 1974. 1977. Conducted by the Ce
Judah Matras and Dov Weintraub, principal investigators. J
of Statistics.
Jackson, John A., John L. P. Thompson, and Donald J. Treii
Histories in the Irish Republic." Paper presented at the meeting
on Stratification of the International Sociological Association, E
Republic, October 29–November 2.
Kahne, Hilda, with Andrew I. Kohen. 1975. "Economic Perspecti
in the American Economy." *Journal of Economic Literature* 1
Kohen, Andrew I. 1977. *Women and the Economy: A Bibliogr*
Literature on Sex Differentiation in the Labor Market. Rev. C
versity, Center for Human Resource Research.
Kosa, John, and Robert E. Coker, Jr. 1971. "The Female PI
Conflict and Reconciliation of the Sex and Professional Rc
Professional Woman, edited by Athena Theodore. Cambridge
Loeb, Jane W., Marianne A. Ferber, and Helen M. Lowry. 19
Affirmative Action for Women." *Journal of Higher Education*

McLaughlin, Steven D. 1975. "Occupational Characteristics and the Male-Female Income Differential." Ph.D. dissertation, Washington State University.

McNulty, D. 1967. "Difference in Pay between Men and Women Workers." *Monthly Labor Review* 90:140–43.

Miller, Herman P. 1966. *Income Distribution in the United States: A 1960 U.S. Bureau of the Census Monograph.* Washington, D.C.: Government Printing Office.

Mincer, Jacob, and Solomon Polachek. 1974. "Family Investments in Human Capital: Earnings of Women." *Journal of Political Economy* 82 (suppl.): S76–S108.

———. 1978. "Women's Earnings Reexamined." *Journal of Human Resources* 13:118–34.

NORC General Social Surveys 1974–77. 1977. Conducted by James A. Davis, Tom W. Smith, and C. Bruce Stephenson, National Opinion Research Center, University of Chicago. Roper ed. Storrs, Conn.: Roper Public Opinion Research Center.

O'Kelley, Charlotte G. 1979. "The Impact of Equal Employment Legislation on Women's Earnings: Limitations of Legislative Solutions to Discrimination in the Economy." *American Journal of Economics and Sociology* 38:419–30.

Oppenheimer, Valerie K. 1968. "The Sex-Labeling of Jobs." *Industrial Relations* 7:219–34.

———. 1970. *The Female Labor Force in the United States: Demographic and Economic Factors Governing Its Growth and Changing Composition.* Westport. Conn.: Greenwood.

———. 1974. "The Life-Cycle Squeeze: The Interaction of Men's Occupational and Family Life Cycles." *Demography* 11:227–45.

Political Action: An Eight-nation Study 1973–76. 1979. Conducted by Zentralarchiv für Empirische Sozialforschung. Cologne: Zentralarchiv für Empirische Sozialforschung.

Quality of Employment Survey 1972–73. 1975. Conducted by Robert P. Quinn, Thomas Mangione, and Stanley Seashore, Survey Research Center, Institute for Social Research, University of Michigan. SSA ed. Ann Arbor: Interuniversity Consortium for Political and Social Research.

Roos, Patricia A. 1981a. "Occupational Segregation in Industrial Society: A Twelve-Nation Comparison of Gender and Marital Differences in Occupational Attainment." Ph.D. dissertation, University of California, Los Angeles.

———. 1981b. "Sex Stratification in the Workplace: Male-Female Differences in Economic Returns to Occupation." *Social Science Research* 10:195–224.

Sandell, Steven H., and David Shapiro. 1978. "The Theory of Human Capital and the Earnings of Women: A Reexamination of the Evidence." *Journal of Human Resources* 13:103–17.

Scandinavian Welfare Survey 1972. 1977. Conducted by the Research Group for Comparative Sociology, University of Helsinki. Helsinki: Research Group for Comparative Sociology.

Smith, Kent W., and M. S. Sasaki. 1979. "Decreasing Multicollinearity: A Method for Models with Multiplicative Functions." *Sociological Methods and Research* 8:35–55.

Swafford, Michael. 1978. "Sex Differences in Soviet Earnings." *American Sociological Review* 43:657–73.

Szalai, Alexander. 1972. *The Use of Time.* The Hague: Mouton.

Takahashi, Nobuko. 1976. "Women's Wages in Japan and the Question of Equal Pay." Pp. 111–28 in *Women Workers and Society: International Perspectives.* Geneva: International Labour Office.

Talbert, Joan, and Christine E. Bose. 1977. "Wage Attainment Processes: The Retail Clerk Case." *American Journal of Sociology* 83:403–24.

Treiman, Donald J. 1977. *Occupational Prestige in Comparative Perspective.* New York: Academic Press.

———. 1979. *Job Evaluation: An Analytic Review.* Washington, D.C.: National Academy of Sciences.

Treiman, Donald J., and Heidi I. Hartmann, eds. 1981. *Women, Work and Wages: Equal Pay for Jobs of Equal Value.* Washington, D.C.: National Academy Press.

Treiman, Donald J., and Jonathan Kelley. 1974. "A Comparative Study of Status Attainment." Grant proposal submitted to the National Institute of Mental Health.

Treiman, Donald J., and Kermit Terrell. 1975a. "The Process of Status Attainment in the United States and Great Britain." *American Journal of Sociology* 81:563–83.

———. 1975b. "Sex and the Process of Status Attainment: A Comparison of Working Women

and Men." *American Sociological Review* 40:174–200.

———. 1975c. "Women, Work, and Wages—Trends in the Female Occupational Structure Since 1940." Pp. 157–99 in *Social Indicator Models,* edited by Kenneth Land and Seymour Spilerman. New York: Russell Sage.

U.S. Bureau of the Census. 1977. *Population Profile of the United States: 1976.* Current Population Reports, ser. P-20, no. 307. Washington D.C.: Government Printing Office.

———. 1980. *Money Income of Familes and Persons in the United States, 1978.* Current Population Reports, ser. P-60, no. 118. Washington, D.C.: Government Printing Office.

U.S. Department of Labor. 1979. *The Earnings Gap between Women and Men.* Washington, D.C.: Government Printing Office.

———. 1980. *Perspectives on Working Women: A Databook.* Washington, D.C.: Government Printing Office.

U.S. Internal Revenue Service. 1974. *Statistics of Income—1972: Individual Income Tax Returns.* Publication 79(1–75). Washington, D.C.: Government Printing Office.

Walker, Kathryn E., and Margaret E. Woods. 1976. *Time Use: A Measure of Household Production of Family Goods and Services.* Washington, D.C.: American Home Economics Association, Center for the Family.

Wallace, Phyllis. 1976. "Impact of Equal Employment Opportunity Laws." Pp. 123–45 in *Women and the American Economy: A Look to the 1980's,* edited by Juanita M. Kreps. Englewood Cliffs, N.J.: Prentice-Hall.

Winsborough, H. H., and Peter Dickinson. 1971. "Components of Negro-White Income Differences." Pp. 6–8 in *Proceedings of the American Statistical Association, Social Statistics Section,* edited by Edwin G. Goldfield. Washington, D.C.: American Statistical Association.

Zumabus 1976. 1976. Conducted by Zentrum für Umfragen Methoden und Analysen. Mannheim: Zentrum für Umfragen Methoden und Analysen.

Social Contexts, Social Networks, and Urban Neighborhoods: Environmental Constraints on Friendship Choice[1]

R. Robert Huckfeldt
University of Notre Dame

This article considers the influence of neighborhood social contexts on the content of social networks. Contextual explanations for individual behavior argue that (1) individual preferences and actions are influenced through social interaction, and (2) social interaction is structured by the social composition of the individual's environment. Thus, a preliminary step to constructing contextual theories of individual behavior is an examination of the way that the social context structures social encounters and friendship choice. The empirical correspondence between the content of neighborhood social contexts and the content of social networks is examined using a sample of white male Detroit residents. A mathematical model of associational choice is developed that incorporates the ability of individuals to enforce social preferences on their choice of friends while it maintains the role of the social context in structuring that choice.

Selecting a friend is among the most personal of human choices, and thus it is not surprising that friendship groups tend toward social class homogeneity. Members of the working class usually associate with other workers, and middle-class individuals generally choose friends who are middle class. But are these friendship choices wholly dependent on the characteristics and preferences of the individual making the choice, or are they also constrained by the availability of socially similar individuals within the environment?

This article relates the social contexts of urban neighborhoods to the social class content of urban social networks and to the process of friendship selection. A mathematical model is constructed to incorporate both individual preference and environmental constraint in a dynamic representation of the process underlying friendship choice. The model is then applied to an analysis of friendship group formation.

The data base for this analysis is the 1966 Detroit Area Study conducted by Edward Laumann and his associates at the University of Michigan. The Detroit survey contains interviews with 985 adult white males, weighted to produce a sample of 1,013. These survey data have been the subject of extensive analysis (Laumann 1973) and reanalysis (Fischer et al. 1977). Complete documentation of study design and questionnaire content is contained in Laumann (1973). I supervised the merger of 1970 census tract data with the survey, thereby producing a data set with measurement at both the individual level (the survey) and the contextual level (the neighborhood measured as a census tract).[2]

DOES THE ENVIRONMENT CONSTRAIN FRIENDSHIP CHOICE?

Many studies of contextual assimilation establish a basic empirical relationship between individual behavior and the social context: individual behavior tends to move in the direction of a surrounding population's social makeup, even when individual characteristics are taken into account.[3] Thus, for example, middle-class people are more likely to support leftist parties if they reside in contexts that are more heavily working class in their aggregate composition. Two assumptions are generally incorporated within the explanations for these kinds of relationships: (1)

[2] This operational definition of the neighborhood social context does not equate a neighborhood with a cohesive community of friends and acquaintances. Such cohesive communities sometimes exist, but they are tangential to the present argument, and they cannot be measured with census tract data. The *neighborhood* as it is conceived here refers to the shared geographic locale of a residential grouping, and the *neighborhood social context* refers to the population composition of the people who live in the neighborhood. This neighborhood definition subsumes the cohesive community but does not depend on it. The definition is designed to capture the inescapable social relations of any geographically based social collectivity. Census tract data for 1970 are used in preference to 1960 data because they are closer in time to the survey. Some tract aggregation was necessary because 1960 tract numbers were recorded for respondents, and because tract boundaries changed between 1960 and 1970. An interpolated measure was not used for two reasons. First, different population bases were used in collecting employment data in 1960 and 1970. Second, a greater degree of tract aggregation is necessary to produce identical boundaries between 1960 and 1970 than is required merely to match 1970 data to 1960 tracts.

[3] These studies constitute a significant body of accumulated research with a modern origin that traces to Blau's (1957) work on structural effects and the work of Davis, Spaeth, and Huson (1961) on compositional effects. This article grew out of a concern with a set of arguments that relate the social context to political behavior (Langton and Rapoport 1975; Segal and Meyer 1974; Sprague 1976; Przeworski 1974; Wright 1977; and Huckfeldt 1979).

people are influenced through social interaction, and (2) social interaction patterns are structured by the social composition of the relevant environment. In the present effort I ignore the first assumption in order to examine the second. To what extent is intimate social interaction, defined as friendship choice, structured by the social context?

Contextual research typically uses spatially bounded population data to obtain a measure of the social context. This means that the social context is made operational as a county, a voting precinct, a census tract, a workplace, or some other population aggregate. The imposition of the boundaries presents several problems related to both the relevance of the boundaries for associational patterns and the exercise of individual discretion over the choice of associates within a particular set of boundaries.

First, the boundaries used to circumscribe a social context are often criticized as being either too large to capture rich internal variation in social worlds or too small to be treated as self-contained environments with the potential for structuring social interaction. Thus, a person's social context is often seen as being individually unique because it is created from the particular set of environments within which that person resides. Next-door neighbors live in the same neighborhood context, but they are employed by different enterprises, worship at different churches, and drink at different taverns. Each of these environments creates its own unique social context, and these neighbors may therefore encounter very different population mixes. As a result, they have vastly different opportunities for friendship formation, regardless of a shared neighborhood environment.

A second objection to assertions of contextually structured patterns of association points to the individual capacity for choosing associates. Even when people share the same set of contexts—neighborhoods, taverns, churches, washeterias—their personal preferences are still enforced on their environment. They are able to choose associates who are attractive socially and agreeable personally, and they are able to avoid individuals who are either socially or personally obnoxious. For example, Berger (1960) shows that middle-class residents of a newly established working-class suburb are resented by working-class residents, isolated from community life, and less likely to increase their level of social interaction in the new environment. Finifter (1974) shows that Republican autoworkers who are employed in a workplace social context that is heavily Democratic choose to associate with other workers who are also Republicans. They enforce their preferences on the context by constructing a friendship group that serves as a "protective environment."

However, a strict emphasis on individual control over associational patterns and social network content produces its own set of problems. The militant contextualist may be in danger of ignoring the effect of

individual preferences on associational patterns and friendship choice, but the focus on personal control may ignore important external con- straints on supply. The available pool of socially similar individuals varies as a function of context, so that the same set of preferences might produce different friendship groups in different environments.

Because this paper considers only one environment, the neighborhood, the resulting task is twofold: (1) to examine the interplay between indi- vidual choice and contextual constraint in the friendship selection process, and (2) to consider the relevance of the neighborhood as a social context that imposes constraints on friendship choice. Choosing a friend is con- ceived as the outcome of a selection process that occurs through time, but we know little regarding the qualitative nature of this process. In- stead, the process is conceived abstractly, in the mathematical language of a dynamic model.

A MATHEMATICAL MODEL OF FRIENDSHIP SELECTION

Because any process occurs through time, a process should be viewed in dynamic terms even when it is observed at a single point in time. Thus, I conceive of the friendship selection process as a search for compatible associates that occurs through time as individuals encounter other indi- viduals in their environment. An analytical distinction is thus drawn between (1) encounters and (2) friendships or associations. I assume that the environment controls the likelihood of encounters within and between social classes, but that individuals exercise discretion in deciding whether to turn an encounter into an association. In other words, a middle-class individual can hardly avoid encountering his working-class neighbor when they are both mowing their lawns, but the middle-class individual *can* decide whether to invite this neighbor to dinner.

People use numerous criteria when they consider the possibility of turn- ing an encounter into a friendship, but this paper considers only one: the social class of the individuals involved in the encounter. Thus, the process being specified is necessarily an abstraction that ignores many important facets of friendship selection—shared interests, compatible personalities, similar political views, and so on. Finally, ignoring these other criteria for associational discretion, I assume that (1) a member of a given social class will not reject association with another member of the same social class on the basis of a social class criterion, but that (2) a member of a given social class might reject a member of a different social class on the basis of such a criterion. As with any simplifying assumption, exceptions abound. For example, the working-class social climber might indeed reject association with other workers on the basis of a social class criterion, but ignoring this kind of exception does not compromise the analysis.

In order to realize the analytic utility of these distinctions and assumptions, they are translated into a mathematical statement that builds on the work of James Coleman (1964, chap. 16). Coleman develops a model to consider the consequences of group size differences, relative to total population size, for the maintenance of norms or beliefs specific to a group (1964, p. 479). A logically similar model is constructed here that relates the neighborhood social context to the construction of social networks and considers the implications of this relationship for the extent of discretion exercised by individuals in choosing intimate associates. The model is written as follows:

$$F_{ij}(k) = S_j(1) + [1 - S_j(1)] \, r_{ij} S_j(2)$$
$$+ [1 - S_j(1)] r_{ij} [1 - S_j(2)] r_{ij} S_j(3) + \ldots \tag{1}$$

where:

$F_{ij}(k) =$ the probability that a member of group i in context j will form a friendship with another member of group i after k opportunities for friendly association;

$S_j(n) =$ the social density of the in-group, conceived as the proportion of social context j that is composed by fellow members of group i at opportunity n for friendly association, or the probability that such an opportunity for association will involve an encounter with a fellow member of group i, assuming random encounters (as opposed to nonrandom associations);

$1 - S_j(n) =$ the social density of the *out-group*, or the probability of an encounter with a member of the out-group in context j at opportunity n for friendly association; and

$r_{ij} =$ the probability that a member of group i in context j will reject association with a nonmember of i, having encountered one at an opportunity for friendly association.

Translating equation (1) into English may be a useful exercise. Suppose that group i denotes the middle class and that all individuals belong to either the middle class or the working class. The probability that a middle-class individual will form an association with another member of the middle class after one opportunity for forming a friendship is equal to the probability of encountering a middle-class individual at the first opportunity. The probability that a middle-class individual will form an association with another member of the middle class after two opportunities for forming associations is equal to (1) the probability of encountering a middle-class individual at the first opportunity plus (2) the probability of encountering a working-class individual at the first opportunity, times the probability of rejection, times the probability of en-

countering a middle-class individual at the second opportunity. This dynamically specified process proceeds indefinitely.

Let me summarize and review the assumptions that are embedded within the model. First, I am assuming a difference between encounters and associations. Encounters occur randomly, and they take place on a daily basis in many different social situations: taking walks in a neighborhood, talking across backyard fences, and so on. Friendships do not occur randomly: they are subject to random encounters, but they are also subject to the preferences of those involved in the associations. Second, in an assumption that will be discussed again, I specify implicitly that a member of a given social class will use a social class criterion to reject association with another member of the same class, given an encounter, with a probability of zero. Third, I am allowing the out-group rejection probability to be specific both to particular groups and to particular contexts. Thus, the rejection probability might vary across groups in a particular context and across contexts for a particular group. Finally, I assume that the social density of a particular group is constant across associational opportunities, and the model may therefore be rewritten as:

$$F_{ij}(k) = S_j + (1 - S_j)r_{ij}S_j + (1 - S_j)^2r_{ij}^2S_j + \ldots$$
$$+ (1 - S_j)^{k-1}r_{ij}^{k-1}S_j, \tag{2}$$

or

$$F_{ij}(k) = \sum_{n=0}^{k-1} S_j(1 - S_j)^n r_{ij}^n. \tag{3}$$

Finally, algebraic manipulation produces the following closed form:[4]

$$F_{ij}(k) = \frac{S_j - S_j r_{ij}^k (1 - S_j)^k}{1 - r_{ij}(1 - S_j)}. \tag{4}$$

As the person becomes settled in a social context and the number of opportunities for social interaction becomes very large, k also becomes very large. And, because r_{ij} and $1 - S_j$ are probabilities, they both lie between zero and one (ignoring the special cases when they equal zero or one), thereby reducing the quantity $[S_j r_{ij}^k (1 - S_j)^k]$ to zero. Thus, the model approaches

---

[4] Algebraic manipulation of an open-form into a closed-form expression is illustrated in Huckfeldt, Kohfeld, and Likens (1982, p. 16).

$$F_{ij} = \frac{S_j}{1 - r_{ij}(1 - S_j)} \text{ as } k \longrightarrow \infty \text{ (as } k \text{ gets large).} \tag{5}$$

Equation (5) is a causal statement based on time-ordered logic. It portrays the dynamic interplay, at equilibrium,[5] between individual discretion and contextual constraint in affecting the process of friendship choice. The discretion parameter r_{ij} deserves more careful scrutiny. If r_{ij} equals zero, F_{ij} equals S_j. That is, if a group member does not resist association with nonmembers, then the probability of associating with a nonmember is equal to the probability of encountering a nonmember. Conversely, when r_{ij} equals one, F_{ij} equals one. This is the pure case of contextual isolation: a group member who rejects all nonmembers as associates will associate exclusively with members, regardless of the social context.

Clearly, r_{ij} is a theoretically abstract concept that cannot be observed directly. It sets forth one dimension of the class or group consciousness held by a particular collectivity. By not incorporating a rejection term for the members of one's own group I ignore the problems posed by individuals whose associational preferences lie beyond the boundaries of their own group. Note, however, that this is a conservative assumption: if I assume that an individual rejects members of his own class with some probability, the out-group rejection parameter must necessarily increase in size to produce the same in-group association probability (F_{ij}) at the same contextual density (S_j).[6]

The model can be applied usefully in analyzing the process of friendship choice among the Detroit respondents. Before proceeding with this analysis, however, I examine the homogeneity and location of respondents' friendship groups.

GEOGRAPHICAL SOURCE AND HOMOGENEITY OF FRIENDSHIP GROUPS

Each respondent to the survey was asked to name his three closest friends. Any relative included on the list was replaced by a nonrelative. These

[5] "Equilibrium" is a technical concept with useful substantive interpretations (Schelling 1978, pp. 25–27). In the present substantive context, it suggests an individual who has resided in an environment for a sufficient amount of time and experienced sufficient social encounters to establish a stable network of friendships.

[6] I am grateful to an anonymous referee for this observation regarding the effect of rejecting in-group associates on the out-group rejection parameter. An additional problem with the out-group rejection parameter relates to the source of rejection. Members of a first group may associate infrequently with members of a second group because the second group will not accept them as associates. This problem is reduced (but not eliminated) through the use of an asymmetrical measure of association: I do not corroborate a respondent's list of friends by examining whether the friendship object reciprocates the friendship. For a discussion of friendship reciprocity and these data, see Laumann (1973).

resulting three choices serve to make the concepts of friendship group and social network operational.[7] My concern is only with the content of social networks, not with their form. As a result I ignore issues of connectedness—the extent to which respondents' friends are associated with each other.

The homogeneity of friendship groups is well-established in social science research, and the friendship groups of the Detroit respondents bear out that homogeneity. Table 1 shows that the working-class respondents and both groups of middle-class respondents (professional-managerial and clerical-service-sales) are socially integrated: that is, their friendships are

TABLE 1

PERCENTAGE OF RESPONDENTS WITH VARIOUS NUMBERS OF FRIENDS IN EACH CLASS
BY RESPONDENT'S OCCUPATIONAL STATUS

| | RESPONDENT'S OCCUPATIONAL STATUS | | | |
	Professional-Managerial	Clerical-Service-Sales	Working Class	ALL
Number of working-class friends:				
3	5.1	9.2	40.0	23.1
2	9.2	18.3	30.8	21.3
1	28.7	31.7	20.0	24.8
0	57.0	40.8	9.2	30.8
Number of clerical-service-sales friends:				
3	1.0	2.8	.4	1.0
2	11.8	25.4	4.3	10.2
1	24.5	40.8	23.6	26.7
0	62.7	31.0	71.7	62.1
Number of professional-managerial friends:				
3	32.2	9.2	2.7	13.9
2	33.8	19.0	13.5	21.4
1	22.6	38.7	29.9	28.6
0	11.5	33.1	53.9	36.0
N	314	142	445	905

SOURCE.—Inter-university Consortium for Political and Social Research (ICPSR) 1977.

[7] Recent investigations of informant accuracy have called into question the quality of social network data: "People simply do not know, with any degree of accuracy, with whom they communicate" (Killworth and Bernard 1976, p. 283). This criticism is probably overstated, at least as it relates to the current effort. Killworth and Bernard examine the accuracy with which informants remember the frequency and duration of communications, but the conception of network used here ("close friends") is based on a personal intensity measure that is necessarily a subjective judgment.

more heavily concentrated within their own occupational group.[8] A member of one group is more likely than a nonmember to associate with other members. Thus, for example, clerical-service-sales respondents are more likely to have clerical-service-sales friends than are either working-class or professional-managerial respondents.

Owing in part to the limited numbers of respondents, I will, later in the analysis, combine the middle-class categories. This practice raises a question: Given the extent of the social fracture between the two groups of middle-class respondents, does the concept of a single middle class have any meaning? The clerical-service-sales respondents occupy an intermediate position not only in terms of individual status but also in terms of friendship patterns. They are more likely than the working-class respondents to have professional-managerial friends, they are more likely than the professional-managerial respondents to have working-class friends, and they are the least introverted socially of the three occupational groups. However, their social ties to the professional-managerial group are clearly stronger. Professional-managerial respondents are more likely to have clerical-service-sales friends than are the working-class respondents. And clerical-service-sales respondents are equally likely to have working-class or professional-managerial friends, even though the working class is more numerous in the population.[9] Finally, approximately two-thirds of the clerical-service-sales respondents identify themselves as middle class (see n. 8 above). Thus, while it is necessary to take account of internal diversity, it is also meaningful to think of a single middle class.

A second characteristic of friendship groups, their geographical location, is displayed in table 2, part A. Respondents were asked whether each of their friends lived (1) in their own neighborhood, defined as a 10-

[8] The measure for social class being used here is designed to separate individuals on the basis of working conditions, job security, prestige or social status, and monetary compensation. Perhaps the best justification for these categories is the respondents' own self-perceptions regarding class membership. Of those who identify with a social class, 78% of the professional-managerial respondents, 68% of the clerical-service-sales respondents, and 35% of the working-class respondents identify themselves as upper or middle class. However, this measure for objectively defined class membership suffers from some obvious shortcomings. Some foremen might be better classified as managers rather than workers, and some service employees might be better classified as workers. For this reason it becomes essential that controls for other individual status characteristics be introduced in later contextual analyses. These controls help to guard against the danger of inferring a contextual relationship that is actually an artifact of unconsidered individual-level factors (Hauser 1974).

[9] According to 1970 census figures, white professional-managerial men constitute approximately 27% of all white employed men in the Detroit SMSA and 23% of all white employed men in the city of Detroit. White working-class men constitute approximately 40% of white employed men in the Detroit SMSA and approximately 52% of white employed men in the city. The sample evidently overestimates professional-managerial presence in the population and underestimates clerical-service-sales presence. Approximately 33% of white employed men in the metropolitan area and approximately 25% of white employed men in the city of Detroit are clerical-service-sales.

minute trip; (2) beyond their neighborhood but within the Detroit area; or (3) beyond the Detroit area.[10] The number of friends living inside the respondent's neighborhood is shown for each of the three occupational categories. Between 36% and 41% of all three groups have a majority of friends living in their neighborhoods, and less than a third of the members of each group do not have any friends in the neighborhood. Finally, the working-class respondents are only slightly more prone to have locally rooted friendship groups.

Thus, table 2, part A, qualifies an analysis of urban neighborhoods that characterizes them as a place for neighborly relations but not as a

TABLE 2

GEOGRAPHICAL LOCATION OF FRIENDSHIP GROUPS

A. PERCENTAGE OF RESPONDENTS WITH VARIOUS NUMBERS OF NEIGHBORHOOD
FRIENDS BY RESPONDENT'S OCCUPATIONAL STATUS

	RESPONDENT'S OCCUPATIONAL STATUS		
NUMBER OF FRIENDS IN NEIGHBORHOOD	Professional-Managerial	Clerical-Service-Sales	Working Class
3	15.2	18.7	19.1
2	23.7	18.0	21.7
1	30.7	32.0	31.5
0	30.4	31.3	27.6
N	329	150	460

B. PERCENTAGE OF RESPONDENTS WITH A MAJORITY OF NEIGHBORHOOD FRIENDS BY
NEIGHBORHOOD PROPORTION OF WORKING CLASS AND BY RESPONDENT'S
OCCUPATIONAL STATUS*

	NEIGHBORHOOD WORKING-CLASS PROPORTION†			
RESPONDENT'S STATUS	Low	Medium-Low	Medium-High	High
Professional-managerial	40.0	36.4	33.3	44.7
	(120)	(77)	(57)	(38)
Clerical-service-sales	36.8	40.0	40.0	36.0
	(38)	(45)	(25)	(25)
Working class	48.2	37.1	40.3	41.7
	(56)	(97)	(129)	(151)

SOURCES.—ICPSR 1977; U.S. Bureau of the Census 1972.

* Numbers in parentheses are N's.

† Cutting points are chosen on the basis of quartile divisions for the sample: low = .058–.448; medium-low = .448–.558; medium-high = .558–.638; high = .638–.823.

[10] The 10-minute transportation definition of neighborhood is adequate for present purposes, corresponding roughly to the census tract definition employed here.

place for forming friendships. The neighboring role may not necessitate close interpersonal association (Keller 1968, pp. 10–11), but it is a mistake to ignore the physical proximity of neighborhood residents as a factor that facilitates their incorporation into a friendship group. This evidence also calls into question a distinction between locally oriented and more cosmopolitan city dwellers, the former being socially anchored in the local area and the latter being oriented toward the city in a way that erases social ties to the neighborhood. An assumption frequently underlying these categories is that lower classes tend to be more local in their orientations (Keller 1968, pp. 116–23), but the evidence in table 2, part A, does not support such an assertion.

Finally, people might be expected to have more local friends in an area where their own social group is more heavily represented. Individual freedom in the choice of associates allows individuals to look beyond the neighborhood if they do not find neighborhood residents to be compatible socially. However, table 2, part B, does not support this expectation. No clear pattern is revealed in the number of local friends across context for any single occupational class.[11]

In summary, tables 1 and 2 have confirmed the social homogeneity of the respondents' friendship groups, but they have also established the importance of physical proximity in friendship group construction. These empirical relationships are suggestive, but they reveal little regarding the process of associational choice or the interplay between individual preference and environmental constraint in affecting that process. In order to consider these issues, I return to the mathematical representation of the process developed earlier and apply it to the Detroit respondents.

APPLYING THE MODEL TO DETROIT RESIDENTS

The model of associational choice, as it is expressed in (5), can be applied usefully to a consideration of the Detroit respondents, but the mathematical language of the model must first be translated into measurable, operational terms. First, the value for F_{ij} is best interpreted as the probability that a single friend will be a member of the respondent's occupational class. Furthermore, in the ensuing analysis I shall draw a distinction between local and nonlocal friends. A locally rooted network might extend beyond the neighborhood so that the local population composition could have consequences for the social class of nonlocal friends,

[11] This analysis is not intended to give detailed consideration to the issue of localism, but it is important for later analyses to show that different occupational groups do not show widely varying degrees of localism and that localism does not greatly vary within a single occupational group across working-class contexts. For a more extended treatment using these data see Fischer et al. (1977, chap. 7).

but the neighborhood effect would be mediated by the point of origin—
the local friend. Thus, in applying the model, I concentrate on respondents
who have at least one local friend and consider the probability that the
first mentioned local friend is a member of the respondent's social class.

As a second step, "in-group" and "out-group" densities must be defined
relative to each of the occupational classes. In order to make the modeling
task manageable, I separate neighborhood social densities into the work-
ing-class proportion and its complement.[12] This complement comes very
close to being an exact measure of the middle-class proportion and it is
treated as such. The in-group density (S_j) for working-class respondents
consists of the neighborhood working-class proportion, and the out-group
density $(1 - S_j)$ consists of the neighborhood middle-class proportion.
Conversely, the in-group density (S_j) for professional-managerial and cler-
ical-service-sales respondents consists of the neighborhood middle-class
proportion, and the out-group density $(1 - S_j)$ consists of the neighbor-
hood working-class proportion. Thus, r_{ij} is the probability that working-
class individuals reject middle-class individuals as associates, or the prob-
ability that middle-class individuals reject working-class individuals, de-
pending on the definition of S_j and $1 - S_j$ and the particular occupational
class of respondents with which I am concerned.

Equation (5) can be rearranged algebraically to express r_{ij} solely as a
function of F_{ij} and S_j:

$$r_{ij} = \frac{F_{ij} - S_j}{F_{ij}(1 - S_j)}. \tag{6}$$

Thus, given a sequence of coupled values for F_{ij} and S_j, I can arrive at
a matched sequence of values for r_{ij}, and the first task is to estimate F_{ij}
as a function of S_j. The logit model for microdata is well-suited to the
task of estimating the probability that the first-mentioned local friend is
a member of the respondent's class (Hanushek and Jackson 1977; McCrate
and Miller 1979). Table 3 shows the logit coefficients that are obtained
when having a local friend who belongs to the respondent's own class is
estimated as a function of context for each social class. In order to com-
pensate for the combination of the two middle-class categories, two ad-
ditional variables are included in the middle-class equation: a dummy
variable for clerical-service-sales status, and an interaction variable formed
by the product of this dummy variable and neighborhood social com-
position. These two variables allow for both individual-level differences

[12] The working-class density is calculated on the basis of employed males in order to obtain
a single, dominant measure for each household in the neighborhood. Thus, a household
with a working-class male and a clerical female would be counted as working class.

and contextual-responsiveness differences between the two middle-class categories. In addition, controls are included for individual income and education and for the respondent's self-perceived class identification.[13] Both the clerical-service-sales dummy variable and its interaction term fail to produce robust t-values. Thus, the middle-class respondents are treated as a single group for the remainder of this analysis.[14]

The coefficients produced by the logit model are useful to the task of estimating the contextually contingent probability of friendly association within the respondent's own class, but they do not estimate the probability as a linear function. Rather, they estimate the "logit"—a nonlinear transformation of the probability. As a result, the magnitude of a coefficient cannot be assessed directly. In order to determine the magnitude of an effect due to an explanatory variable, predicted probabilities are calculated across the range of that variable with all other variables held constant.[15]

The logit coefficients of table 3 are used to estimate F_{ij}—the contextually contingent probability that respondents from each social class will have a friend who is from the same social class—while holding all other individual status measures constant. In order to draw a direct comparison between the two social classes, figure 1A shows the probabilities that a local friend is working class, across neighborhood contexts for respondents from each social class. For both classes, the probability that a local friend is working class increases significantly as a function of increased working-

[13] As suggested earlier, individual controls are introduced in an effort to satisfy one of Hauser's (1974) well-taken objections to assertions of contextual influence. These individual-level factors are likely to covary with both the social context and associational preference, and excluding them might give weight artificially to a contextual interpretation. Even these precautions do not wholly resolve the problem of working-class individuals who choose middle-class neighborhoods because they already have middle-class friends.

[14] I combine the 89 clerical-service-sales respondents with 196 professional-managerial respondents to obtain a larger sample. Such a procedure is substantively appropriate if I can justify the estimation of a main effect for the social context among the middle class as a whole. Analyzing these clerical-service-sales respondents separately produces a coefficient for S that lies in the appropriate direction but a t-value that fails to achieve a robust magnitude (.88). Thus, an alternative interpretation is that the neighborhood effect has not been demonstrated for clerical-service-sales respondents. Readers who adopt this latter interpretation should view the table 3 and fig. 1 middle-class results as being valid only for the 196 professional-managerial respondents. A robust t-value persists for professional-managerial respondents even when they are analyzed separately (3.72).

[15] The logit is equal to the natural log of the odds ratio and is expressed as a linear function of the explanatory variables: $\ln(P/1 - P) = f$, where P is the probability and f is the linear function of explanatory variables and their coefficients. Isolating P on the left-hand side produces $P = 1/(1 + e^{-f})$. The first partial derivative of this function with respect to any explanatory variable does not isolate that variable's coefficient but produces instead an expression that includes the coefficient as well as all other explanatory variables and their coefficients. Thus, a single coefficient must always be judged relative to all other explanatory variables and their coefficients, or, in other words, to a particular point on the probability distribution.

class densities, even though pronounced individual-level differences in the probability exist between the occupational classes.[16]

The out-group rejection parameter is shown specific to particular working-class contexts for each occupational class in figure 1B. Combining the two parts of figure 1 produces a startling picture. In spite of the

TABLE 3

HAVING A LOCAL FRIEND WHO BELONGS
TO SAME SOCIAL CLASS

	RESPONDENT'S SOCIAL CLASS	
	Middle Class*	Working Class
a_1	−2.88	−.40
	(3.48)	(.41)
a_2	.18	−.19
	(1.84)	(1.75)
a_3	.07	−.06
	(.51)	(.46)
a_4	.58	−.24
	(1.77)	(.87)
a_5	.81	. . .
	(.70)	
a_6	−1.36	. . .
	(.58)	
a_7	4.74	3.59
	(3.52)	(3.20)
N	285	313

NOTE.—As a function of individual income and education, self-perceived social class identification, and neighborhood working-class density, for each occupational group. The t-value is shown in parentheses. $P = 1/(1 + e^{-f}); f = a_1 + a_2E + a_3I + a_4ID + a_5CSS + a_6INT + a_7S$. $P = 1$ if first mentioned local friend belongs to respondent's social class, 0 otherwise. E = respondent's education: 1 = 0–8 grades, 2 = some high school, 3 = high school graduate or nongraduate with vocational training, 4 = high school graduate with vocational training, 5 = some college, 6 = college graduate, 7 = college graduate with graduate training. I = respondent's income: 1 = under $3,000, 2 = $3,000–$4,999, 3 = $5,000–$6,999, 4 = $7,000–$9,999, 5 = $10,000–$14,999, 6 = $15,000–$19,999, 7 = $20,000–$24,999, 8 = $25,000 and over. $ID = 1$ if respondent identifies as upper or middle class, 0 if respondent identifies as working class. $CSS = 1$ if respondent is clerical-service-sales, 0 otherwise. INT = the product of in-group density and clerical-service-sales dummy variable: $(CSS) (S)$. S = proportion of employed neighborhood men who belong to the same social class as the respondent.

* The middle-class category combines clerical-service-sales and professional-managerial respondents.

[16] The contextual effect on friendship choice persists, even if the composition of the entire friendship group is considered as the dependent variable in the equation (Huckfeldt 1982). As we might expect, however, this alternative analysis produces a neighborhood effect that is somewhat smaller in magnitude.

664

relatively high rates of change in the probabilities of having a working-class friend, we witness some very high levels of resistance to association across class boundaries. The predicted probability that a working-class respondent has a working-class friend increases significantly across the observed range of working-class densities, from .24 in a 6% working-class neighborhood to .83 in an 82% working-class neighborhood. The out-group rejection parameter decreases across that same range, from .80 at the lowest working-class density to .08 at the highest working-class density. A greater supply of working-class residents in the neighborhood also produces an increase in the probability of working-class association for middle-class respondents, from .04 at the lowest working-class density to .62 at the highest. For these middle-class respondents, however, the out-group rejection parameter increases overall, from .30 at the lowest working-class density to .64 at the highest. Thus, for both working-class and middle-class respondents, the tendency to reject friendships across class lines is generally higher when (1) the social density of the opposite class is higher, and (2) the probability of friendship with a member of the opposite class is also higher! Regardless of strong associational preferences, the formation of friendships is circumscribed by the opportunities and constraints of context. Individual preference is overwhelmed by factors related to supply.

Figure 1B can be compared usefully with table 2B: while minority status within a neighborhood does not necessarily force individuals to look elsewhere for friends, it is a mistake to ignore the forceful expression of social class preference in the formation of friendship groups. The clear implication is that people in a minority must look more diligently within the context for socially similar associates, and their search is often un-
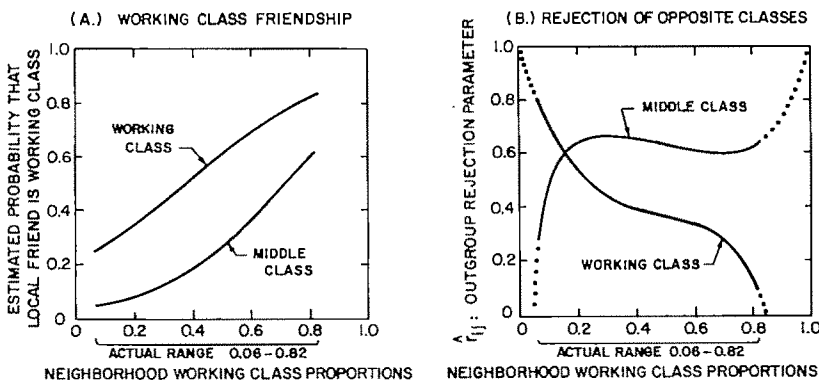
FIG. 1.—Working-class friendship and rejection of opposite classes, across working-class contexts. (These results are based on the logit model of table 3. Individual income and education are controlled at the sample means; subjective social class is controlled at the objectively defined class membership.)

successful. Furthermore, the process specified in equation (2) takes more time to reach equilibrium because the search for associates must be extended. I have assumed that respondents' friendship groups are at equilibrium, but in this sense the minority is more likely to be at disequilibrium: unstable socially and divorced from any strong social ties within the context. At the very least, this kind of social equilibrium will take longer to achieve for members of the minority.

Finally, figure 1B also shows that resistance to the out-group varies as a function of context. As an occupational class approaches minority status, the probability generally increases that members will reject association with nonmembers. Equation (6) shows that when S_j equals zero, r_{ij} must equal one. In substantive terms, however, no one will be available to do the rejecting. When S_j equals one, r_{ij} is mathematically undefined: in substantive terms, there is no one available to reject. Between these extreme values, any value for r_{ij} that lies between zero and one is mathematically possible. For example, an out-group rejection parameter of .01 is possible at an in-group density of .01, but it would produce an in-group friendship probability (F_{ij}) of only .0101. Our results, and the logic of the model, suggest that people resist this sort of social domination by increasing their level of resistance toward out-group friendships. Conversely, an out-group rejection parameter of .99 would be possible at an in-group density of .99, producing an in-group friendship probability of .9999. But, if r_{ij} is reduced to .1, the in-group rejection probability would still equal .9901. The logic of the model suggests that the additional investment in a higher level of resistance (r_{ij}) is simply not worth the effort, and the empirical results indicate that people tend to relax their resistance toward the out-group when their own group is ascendant. Thus, to the extent that the model portrays the search for friends adequately, groups turn inward naturally when they occupy a more pronounced minority status. As a result, F_{ij} exceeds S_j considerably at low in-group densities, but the two values are more nearly equal at high in-group densities.[17]

Empirical documentation of this inward search for associates on the part of minority-status social groups is presented in a number of studies. For example, the previously discussed work of Finifter (1974) demon-

---

[17] Note that F_{ij} must be greater than S_j for r_{ij} to be a believable probability. This is merely a restatement of an earlier assumption: the probability of associating with an in-group member must at least equal the probability of encountering an in-group member, or the probability of rejecting an in-group member equals zero. Negative values for r_{ij}, produced by values of F_{ij} less than S_j, signify a breakdown in the model. Such a situation indicates that in-groups and out-groups are defined improperly, at least for that particular context, or that friendship selection has become socially indiscriminant in these contexts. Projecting the values for r_{ij} beyond the observed range of working-class densities in figure 1B produces negative values.

666

strates a turn inward on the part of minority-status Republican auto-workers but associational openness on the part of majority status Democratic autoworkers. Berger's (1960) study of a working-class suburb documents the turn inward and residential relocation of upper-status residents. And Gans (1967) shows a turn inward on the part of those people who did not "fit in" at Levittown: Jews, older people, the working class, and so forth. The qualitative nature of this inward turn differs across groups and environments, but the underlying logic is the same.

SUMMARY AND CONCLUSION

Evidence presented here shows that the social class content of friendship groups is influenced by associational opportunities and constraints imposed by the neighborhood social context. A person residing in a social context where a particular class is more dominant is more likely to have a friend from that class, regardless of his own class membership.

The interplay between associational preference and the social context in producing a social network is probably more important than this relationship between contexts and networks. Even though individuals demonstrate strong associational preferences, their contextually structured set of associational opportunities makes itself felt in the composition of friendship groups. Thus, the social content of social networks is not solely a function of either the social context or individual choice; it is the complex product of individual preferences operating within the boundaries of a social context.

This effort has constructed a demanding test for the efficacy of social contexts in structuring social interaction. I consider only the role that neighborhood contexts play in altering the construction of intimate friendship groups, but the social context is likely to affect other interactions as well (Sprague 1982). In particular, the content of social encounters (as opposed to associations or friendships) is likely to be much more responsive to the social composition of an environment. By showing that the context structures the formation of friendship groups, this analysis does not encapsulate contextual influence, but it does show that even intimate social interactions are structured by associational opportunities and constraints that are imposed contextually. Theories of contextual influence must work toward carefully constructed typologies of social interaction, empirically documented relationships between the social context and these different types of interaction, and explanations for behavior that rely on these relationships.

Finally, this article suggests several profitable lines of inquiry related to the influence of social contexts. Theories of behavior relying on contextual influence would do well to acknowledge the nonrandom associ-

ational patterns underlying such influence. If contextual influence operates through the formation of social networks, statements regarding the connections between contexts, networks, and behavior should be specified more fully. Contextual influence is not simply a matter of assimilation and absorption. These friendship patterns also point to processes of exclusion, rejection, and hostility; but such responses to the social context are likely to be ignored as long as social interaction within social contexts is treated in a primitive manner.

REFERENCES

Berger, Bennett M. 1960. *Working-Class Suburb: A Study of Auto Workers in Suburbia.* Berkeley and Los Angeles: University of California Press.
Blau, Peter M. 1957. "Formal Organizations: Dimensions of Analysis." *American Journal of Sociology* 58:58–69.
Coleman, James S. 1964. *Introduction to Mathematical Sociology.* New York: Free Press.
Davis, James. A., Joe L. Spaeth, and Carolyn Huson. 1961. "Analyzing Effects of Group Composition." *American Sociological Review* 26:215–25.
Finifter, Ada W. 1974. "The Friendship Group as a Protective Environment for Political Deviants." *American Political Science Review* 68:607–25.
Fischer, Claude S., Robert Max Jackson, C. Anne Stueve, Kathleen Gerson, Lynne McCallister Jones, with Mark Baldassare. 1977. *Networks and Places: Social Relations in the Urban Setting.* New York: Free Press.
Gans, Herbert J. 1967. *The Levittowners.* New York: Vintage.
Hanushek, Eric A., and John E. Jackson. 1977. *Statistical Methods for Social Scientists.* New York: Academic Press.
Hauser, Robert M. 1974. "Contextual Analysis Revisited." *Sociological Methods and Research* 2:365–75.
Huckfeldt, R. Robert. 1979. "Political Participation and the Neighborhood Social Context." *American Journal of Political Science* 23:579–92.
——— 1982. "Political Loyalties and Social Class Ties: The Mechanisms of Contextual Influence." Working Paper. Notre Dame, Ind.: Department of Government, University of Notre Dame.
Huckfeldt, R. Robert, C. W. Kohfeld, and Thomas W. Likens. 1982. *Dynamic Modeling: An Introduction.* Sage University Paper Series on Quantitative Applications in the Social Sciences, no. 07-001. Beverly Hills, Calif.: Sage.
Inter-university Consortium for Political and Social Research (ICPSR). 1977. *1966 Detroit Area Study.* Ann Arbor: ICPSR.
Keller, Suzanne. 1968. *The Urban Neighborhood: A Sociological Perspective.* New York: Random House.
Killworth, Peter D., and H. Russell Bernard. 1976. "Informant Accuracy in Social Network Data." *Human Organization* 35:269–86.
Langton, Kenneth P., and Ronald Rapoport. 1975. "Social Structure, Social Context, and Partisan Mobilization: Urban Workers in Chile." *Comparative Political Studies* 8:318–44.
Laumann, Edward O. 1973. *Bonds of Pluralism: The Form and Substance of Urban Social Networks.* New York: Wiley.
McCrate, M. Martha, and J. Phillip Miller. 1979. "The Predict Procedure." Pp. 75–83 in *SAS Supplementary Users' Guide,* edited by Jane T. Helwig. Raleigh, N.C.: SAS Institute.
Przeworski, Adam. 1974. "Contextual Models of Political Behavior." *Political Methodology* 1:27–61.
Putnam, Robert D. 1966. "Political Attitudes and the Local Community." *American Political Science Review* 60:640–54.

Schelling, Thomas C. 1978. *Micromotives and Macrobehavior.* New York: Norton.

Segal, David R., and Marshall W. Meyer. 1974. "The Social Context of Political Partisanship." Pp. 217–32 in *Social Ecology,* edited by Mattei Dogan and Stein Rokkan. Cambridge, Mass.: MIT Press.

Sprague, John. 1976. "Estimating a Boudon Type Contextual Model: Some Practical and Theoretical Problems of Measurement." *Political Methodology* 3:333–53.

———. 1982. "Is There a Micro-Theory Consistent with Contextual Analysis?" Pp. 99–121 in *The Nature of Political Inquiry,* edited by Elinor Ostrom. Beverly Hills, Calif.: Sage.

U.S. Bureau of the Census. 1972. *Census of Population 1970.* Washington, D.C.: Government Printing Office.

Wright, Gerald C. 1977. "Contextual Models of Electoral Behavior: The Southern Wallace Vote." *American Political Science Review* 71:497–508.

Distress and the Traditional Female Role: A Comparison of Mexicans and Anglos[1]

Catherine E. Ross, John Mirowsky, and Patricia Ulbrich
University of Illinois at Urbana-Champaign

Among Anglos, married women have higher levels of psychological distress than married men. One explanation is the role-stress theory developed by Gove and his colleagues, which focuses on the stressful aspects of traditional female roles (such as housewife and child rearer) in a society that does not value women's traditional role in the home and does not provide an effective set of institutions for dealing with its demands The implications of this theory for cross-cultural research are clear yet, to date, untested. One implication is that in societies that place more value on the family and on the woman's role in the home, the psychological distress levels of married men and women will be more similar. This implication of Gove's theory is tested in two cultures, one that places relatively more emphasis on individual achievement outside the home (Anglo culture) and one that places relatively more emphasis on the family and home (Mexican culture). As expected, the gap in psychological distress levels between married men and women is found to be much smaller in Mexican culture than in Anglo culture. However, the explanations for this finding are not as clear-cut as the role-stress arguments imply.

In American society married women have higher rates of psychological distress—including anxiety, depression, worry, and demoralization—than do married men (Gove and Tudor 1973; Pearlin 1975; Radloff 1975). A leading explanation for this finding is the role-stress theory developed by Gove and his colleagues (Gove and Tudor 1973; Gove and Geerken 1977). The theory focuses on the stressful aspects of female roles in modern industrial society, arguing that women whose roles are least traditional

(i.e., most like men's) are least distressed. In particular, married women who are employed, do not have children at home, or do not have primary responsibility for the housework are expected to be less distressed than married women who stay home taking care of the children and doing household chores. Employed women are expected to be less psychologically distressed than housewives because they have two sources of potential gratification (work and family) instead of one; they have more prestige, power, and personal economic resources than housewives; they are less socially isolated than housewives; and they receive more gratification from their occupational roles than housewives receive from housework. Women with children at home (especially young children) are expected to be more psychologically distressed than those without, because children place demands and burdens on mothers and isolate them from adult social support. Women who have primary responsibility for the housework and cooking are expected to be more distressed than women whose husbands share these chores with them because housework is low-prestige, invisible, ungratifying work for which one receives few rewards. Furthermore, employed women who also have primary responsibility for the home may experience work overload. In sum, Gove's theory proposes that traditional female roles—including being a housewife who is not employed outside the home, having the main responsibility for tasks inside the home, and raising children—are expected to produce psychological distress in married women. Given that, in recent history, most married women have occupied traditional female roles, this is a reasonable explanation for the finding that married women are more psychologically distressed than married men. This theory, however, neglects the contribution of marital satisfaction to psychological well-being. Support, intimacy, and equity in a marriage decrease psychological distress, whereas conflict between husband and wife increases distress (Burke and Weir 1976; Schafer and Keith 1980; Vanfossen 1981; Pearlin 1975; Weissman and Paykel 1974; Radloff 1975). Since women tend to be more dissatisfied with their marriages than men (Renne 1970), marital dissatisfaction may partly explain why wives are more distressed than husbands. In sum, the stressful roles occupied by married women in modern industrial societies, coupled with lower levels of marital satisfaction, lead to higher rates of psychological distress.

Although traditional female roles are stressful for women in modern industrial societies, they are not necessarily stressful for women in more traditional societies. According to Gove and Tudor, "the difficulties confronting women [are] a result of recent changes in the woman's role in industrial societies. According to this argument, women previously had a more meaningful role. Families were large, . . . housework . . . was highly valued . . . [and] the wife played a role in supporting the family."

With the development of industrialization and the small nuclear family, women's traditional role in the home became less important and less valued. At the same time the ideological structure changed, emphasizing women's chance for individual achievement and equality. "However, instead of being treated as equals, women remained in their old institutionalized [and now little-valued] positions" (1973, p. 816). The implications of this theory for cross-cultural research are clear yet, to date, untested. In societies that are more traditional in the sense that the family and the woman's role in the home are accorded more value, and individual achievement is accorded somewhat less value, the psychological distress levels of married men and women will be more similar.

We compare persons of Mexican and Anglo cultural identity living in both the United States and Mexico. Different norms and values about work and about the family in Mexican and Anglo culture imply that the mechanisms that produce psychological distress are different in the two cultural groups. To persons of Mexican heritage the family is a very strong and important part of life (Farris and Glenn 1976; Mirandé 1977; Murillo 1976; Lomnitz 1977). People of Mexican heritage receive a great deal of social support from the family and, whereas Anglos derive most of their status from their job, Mexicans may derive more status from their families. This implies a number of modifications of Gove's role-stress hypotheses for Mexicans.

First, the occupation of the Mexican male does not define family status to as great an extent as the occupation of the Anglo male does. In addition, the Mexican female, as the center of the family, is accorded honor and prestige that is not available to her American counterpart (Elmendorf 1977; Baca Zinn 1975). Thus, we hypothesize that although women are more psychologically distressed than men in both groups, the gap between the sexes is less pronounced in Mexican culture than in Anglo culture.

Second, since prestige, power, and self-esteem are not derived from achieved status to as great an extent among Mexicans, employment is more crucial to psychological well-being for Anglo women than for Mexican. The problem for Anglo women is that they live in a society that does not value the housewife role, whereas employment gives them access to increased prestige and self-esteem. In Mexican culture, the role of housewife may be more highly valued and social arrangements may provide more support (Elmendorf 1977). Thus we hypothesize that working outside the home (as compared with being a housewife) decreases psychological distress for Anglo women but has less effect on Mexicans.

Third, while Anglo women are physically and psychologically isolated in the nuclear family, Mexicans are embedded in a larger family network including both extended family and pseudofamily ties (Mirandé 1977). Although nuclear families of Mexican heritage maintain separate house-

holds, they try to locate each household near others in the extended family, thus facilitating mutual psychological and economic support (Keefe, Padilla, and Carlos 1979; Lomnitz 1977; Madsen 1964). This may make Mexicans less dependent on the quality of their marriage for their psychological well-being, since the larger family network provides affection and emotional support. Therefore, we hypothesize that satisfaction with marriage has a larger effect on psychological distress for Anglos than for Mexicans.

Fourth, the support networks available to the Mexican woman, coupled with the importance of the mother's role (Baca Zinn 1975; Elmendorf 1977), imply that the presence of children is less psychologically distressing for Mexican wives than for Anglos.

SAMPLE

Data were collected by means of a survey questionnaire administered in face-to-face interviews in 1975 in El Paso, Texas, and Juarez, Mexico, companion cities on opposite sides of the border separating Mexico and the United States. Blacks, Orientals, American Indians, Jews, and persons not raised in the United States or Mexico were excluded from the survey. In El Paso there are 330 respondents (a response rate of 73%). In Juarez there are 133 respondents (a response rate of 75%). The total number of cases is 463. Since this is a study of married people, we have selected only married respondents, for an N of 326. The questionnaire was administered in Spanish or English, depending on the respondent's preference. (For more information on the sample see Mirowsky and Ross [1983].)

VARIABLES

We are interested in the additive and interactive effects of sex, Mexican ethnic identity, family roles, wife's employment status, children, and marital satisfaction on psychological distress. We also include three measures of socioeconomic status (education, occupational prestige, and income) known to have effects on psychological well-being. Finally, we control for socially desirable response style in order to make sure that sex differences in psychological distress are not simply due to reporting differences.

Mexican ethnic identity is measured by five indicators: self-reported ethnic group (Anglo, Mexican-American, Mexican), whether the interview was in Spanish (the respondent's choice), and the frequency with which Spanish is used in the family, in visiting with friends, and at social gatherings (never, sometimes, usually, or always). This produces an index

673

of Mexican ethnic identity in which, for example, Mexican-Americans who speak English with their friends and family and identify with American culture may score on the non-Mexican end of the continuum, while Mexican-Americans who identify strongly with Mexican culture and speak Spanish with their friends and family may score on the Mexican end of the continuum. Since the scores are bimodal, with most respondents identifying clearly with either Mexican culture or American culture, we dichotomized the index into Mexican (1) versus American (0) identity.

We measure two aspects of family roles: who has the main responsibility for the housework, and who has the main responsibility for the cooking. If the woman has the sole responsibility for the housework, or if the children, mother, or sister help her with the work, housework roles are coded as traditional (1). If the woman shares the responsibility with her husband, or if the husband has the main responsibility for the housework, it is coded as nontraditional (0). Responsibility for cooking is coded in the same way.

Marital satisfaction is an index composed of the responses to three questions: (1) how satisfied are you with your marriage? (coded from very dissatisfied to very satisfied), (2) would you say your marriage is a happy one? (coded from very unhappy to very happy), and (3) would you like to change your marriage or have it continue the way it is going now? (coded from change many parts of it to continue the same way). The three components all load over .5 on a single factor that had an eigenvalue of greater than one. The resulting index is coded from low to high satisfaction (0–10).

Education is measured by years of formal education. Occupational prestige is measured by Duncan SEI scores. Family income is measured in American dollars. Pesos are converted to dollars using the exchange rate in effect at the time of the survey. Family income includes self-reported earnings, bonuses, tips, commissions, public assistance, welfare, social security, interest, dividends, property rental and sale, pensions, and child support.

Wife's employment status contrasts full- or part-time employment (1) with no employment outside the home (0). The number of children is the number under age 21 reported living at home. Socially desirable response style is measured with the Marlowe-Crowne scale (Crowne and Marlowe 1960). Sex is coded 0 = male, 1 = female.

Psychological distress is measured with eight items from the Langner Index (1962) that indicate psychological distress and are not confounded with physical symptoms (Wheaton 1982).[2] The alpha reliability of the

[2] *Psychological distress* refers to unpleasant moods such as worry, anxiety, demoralization, hopelessness, and depression. It does not refer to disordered cognitive or perceptual processes, withdrawn or inappropriate behavior, or psychophysiological malaise.

674

psychological distress index is .79 for the Anglos and .75 for the Mexicans. It is coded from 0 to 4.

METHOD

We first present the mean levels of all variables in the two cultural groups. Next we present a regression analysis with psychological distress as the dependent variable. In addition to examining the additive effects of the independent variables, we test all the following two- and three-way interactions: Mexican identity by sex, wife's employment status by sex and Mexican identity, number of children by sex and Mexican identity, and marital satisfaction by sex and Mexican identity. This allows us to determine whether the effect of sex on psychological distress is significantly different in the two cultural groups, and whether the effects of wives' employment status, children, and marital satisfaction are different for Mexican and Anglo men and women. We also test all the intervactions substituting income for Mexican identity in order to make sure that apparent cultural differences are not actually due to social class. Finally, we develop a causal model of psychological distress. This allows us to examine the intervening mechanisms through which culture and gender affect distress.

RESULTS

Descriptive Statistics

Most married women occupy traditional roles: 70% of the women are not employed outside the home, 96% occupy traditional housework roles, and 96% occupy traditional cooking roles. As Gove and Tudor (1973) pointed out, even if a married woman is employed, she usually has primary responsibility for the home. Of the employed women, 88% have primary responsibility for the cooking. Among women who are not employed, 99% have primary responsibility for the cooking. Of the women who are employed, 92% have primary responsibility for the housework, compared with 97% of the nonemployed women. Although employed women are more likely than housewives to be in a situation where the husband shares housework and cooking with them, the large majority are still responsible for the home.

Mexican and Anglo families differ significantly in the percentage of wives who are employed: 39% of the Anglo wives are employed, while only 21% of the Mexican wives are employed. This difference is significant at $P < .001$. However, Mexican and Anglo families do not differ significantly in the percentage of families with traditional housework and cooking arrangements. Among the Mexican women, 97% occupy tradi-

tional housework and cooking roles, compared with 95% of the Anglos. Although Anglo wives are significantly more likely to be employed than their Mexican counterparts, Anglo husbands are not significantly more likely to share the household tasks than Mexican husbands. These results are interesting because they suggest that most married women—whether Mexican or Anglo, employed or unemployed—occupy traditional roles in the home. Unfortunately, the lack of variance in family roles (only 4% of the women occupy either nontraditional housework roles or nontraditional cooking roles) means that these will not be useful independent variables in the regression analyses, since variables without much variance are unlikely to explain anything.

Table 1 shows the breakdown of individual characteristics by ethnicity and sex. It shows that women in both cultures have fewer years of formal education than the men. The occupational prestige scores, on the other hand, tell a different story. Among the Anglos, women who work have slightly lower occupational prestige than the men, whereas among the Mexicans, the women who are employed have higher occupational prestige than the men. The Mexican men have the lowest-status jobs of any group.

Regression Analysis and Path Model

The first analysis in table 2 shows the significant direct effects on psychological distress. People who are satisfied with their marriages are less

TABLE 1

DESCRIPTIVE STATISTICS BROKEN DOWN BY ETHNICITY FOR FAMILY CHARACTERISTICS AND BY ETHNICITY AND SEX FOR INDIVIDUAL CHARACTERISTICS

	ANGLO		MEXICAN	
	Men	Women	Men	Women
Education*	12.47		6.08	
	13.34	11.81	6.40	5.86
Occupational prestige*...........	519.34		299.99	
	529.33	502.29	274.25	361.52
Marital satisfaction*.............	8.96		8.29	
	9.24	8.75	8.78	7.79
Psychological distress............	.80		.86	
	.58	.97	.77	.92
Family income*.................	16,188		6,415	
N children*	1.88		3.68	
Wife employed (%)..............	39.26		20.75	
Traditional housework roles (%) ...	95.09		96.86	
Traditional cooking roles (%)	95.09		97.48	

NOTE.—Means shown for ordinal and interval-level variables; percentages shown for dichotomous variables.
* Mexicans are significantly different from Anglos, $P < .05$.

676

psychologically distressed than those who are dissatisfied, people with high income are less distressed than the poor, and women are more psychologically distressed than men. When the wife is employed, psychological distress is lower for both husband and wife. (The interaction of sex by wife's employment is insignificant.) Occupational prestige, edu-

TABLE 2

REGRESSIONS FOR RESTRICTED PATH MODEL

INDEPENDENT VARIABLES	DEPENDENT VARIABLE		
	b	Standard Error of b	β
Psychological Distress			
Marital satisfaction	−.082	(.017)	−.255
Income*	−.007	(.003)	−.121
Wife employed	−.120	(.070)	−.090
Mexican	.099	(.107)	.080
Sex (female)	.310	(.090)	.249
Mexsex interaction	−.215	(.128)	−.159
Socially desirable response	−.058	(.013)	−.242
Constant	2.104	. . .	. . .
R^2	.205	. . .	. . .
Marital Satisfaction			
Wife employed	−.629	(.229)	−.152
Education	.096	(.029)	.249
Sex (female)	−.524	(.209)	−.136
Socially desirable response	.121	(.040)	.164
Constant	6.929	. . .	. . .
R^2	.107	. . .	. . .
Income*			
Education	.661	(.146)	.292
Mexican	−5.512	(1.448)	−.246
Constant	7.920	. . .	. . .
R^2	.107	. . .	. . .
Wife Employed			
Education	.022	(.005)	.240
Constant	.096	. . .	. . .
R^2	.058	. . .	. . .
Education			
Mexican	−6.357	(.425)	−.641
Sex (female)	−1.018	(.431)	−.101
Constant	13.018	. . .	. . .
R^2	.424	. . .	. . .

NOTE.—All variables are significant at $P < .05$, one-tailed test, except for the additive effect of Mexican identity on distress, which is included in the equation because it is a component of the interaction term. $N = 326$.

* In thousands of dollars.

677

cation, and the number of children do not have significant direct effects on psychological distress. Only one interaction term is significant—the interaction of sex with Mexican heritage has a significant effect on psychological distress, which can be interpreted in the following way. Being a woman increases psychological distress in both cultures. However, the gap between men and women in psychological distress levels is narrower among Mexicans than among Anglos. Figure 1 is a graphical representation of the interactive effect of sex and Mexican identity on distress, controlling for other variables with direct effects.

In order to examine the indirect effects on psychological distress, we develop a causal model in which employment status and marital satisfaction may be intervening mechanisms, explaining the effects of sex on distress. More specifically, sex and Mexican heritage are exogenous variables. Psychological distress is the dependent variable. Education, the wife's employment status, family income, and marital satisfaction are intervening variables. We also control for socially desirable response style. Given our causal order assumptions,[3] all possible paths were tested. Those
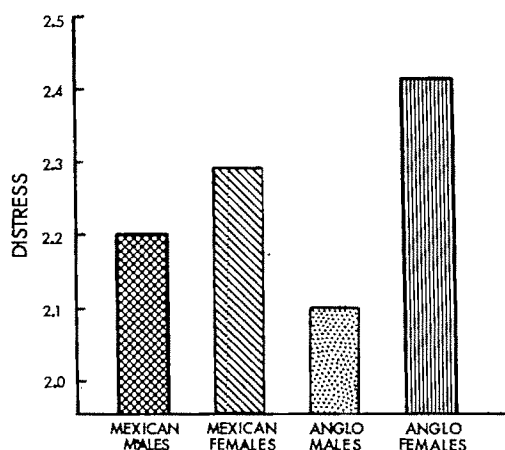
FIG. 1.—Sex and cultural differences in psychological distress controlling for income, wife's employment status, marital satisfaction, and socially desirable response. Psychological distress = 2.10 + .099 (Mexican) + .310 (sex) − .215(Mexican × sex), where Mexican = 1, Anglo = 0; female = 1, male = 0.

[3] The assumption that marital satisfaction affects psychological distress rather than vice versa may be considered problematic. However, there is evidence from other research that marital characteristics often precede psychological distress. Vanfossen (1981) found that 83% of her sample reported that poor marital situations had been going on longer than depressive symptoms, and Brown and Harris (1978) found that intimacy with one's husband prior to the onset of stressful life events protected women against becoming depressed. If there is any effect of psychological distress on marital satisfaction we assume it is small relative to the effect of marital satisfaction on distress. However, we cannot test this with cross-sectional data. We hope future research examines these issues using longitudinal data.

that were not significant at $P < .05$ were deleted, and the model was reestimated with only significant paths included. The final model, with standardized regression coefficients, is shown in figure 2. Table 2 shows the regression analyses for figure 2.

Figure 2 shows that education and the wife's participation in the labor force affect marital satisfaction, which in turn affects psychological distress. Education has the largest direct effect on satisfaction, with more educated people being more satisfied with their marriages ($\beta = .25$). Interestingly, education increases marital satisfaction, but income does not. If the wife is employed, both the wife and the husband are less satisfied with their marriage than when the wife does not work ($\beta = -.15$). This effect may be contrasted with that of the wife's employment on distress: in dual-worker families both the husband and the wife are less psychologically distressed. Thus, the wife's employment decreases distress directly ($-.09$) but increases distress indirectly by decreasing marital satisfaction ($-.15 \times -.26 = .04$). This, along with the findings that income increases psychological well-being but not marital satisfaction .and that education increases satisfaction but has no direct effect on psychological well-being, indicates that the socioeconomic and interpersonal dynamics that produce marital satisfaction differ from those that produce psychological well-being.

Sex, however, has a consistent effect on both marital satisfaction and psychological distress. Married women are less satisfied with their marriages and are more psychologically distressed than married men. Thus, female gender works to increase distress not only directly but also indirectly through dissatisfaction with marriage.[4]
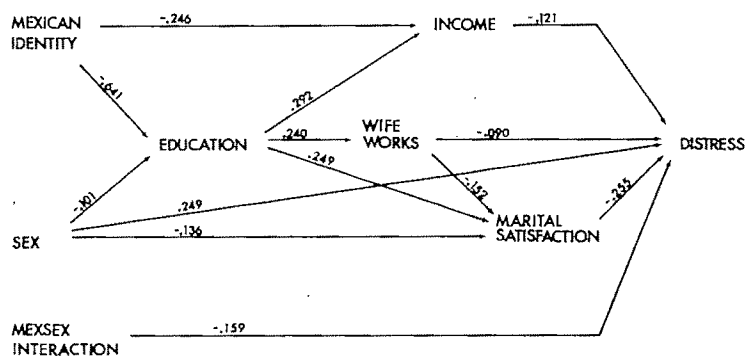
FIG. 2.—Restricted path model of psychological distress (standardized coefficients). All paths are significant at $P < .05$, one-tailed test. Female $= 1$, male $= 0$; Mexican $= 1$, Anglo $= 0$.

[4] Since we have controlled for socially desirable response style, these findings are not likely to be an artifact of reporting.

DISCUSSION.

Gove's basic proposition and the first implication of his theory for cross-cultural research are supported. Married women are more psychologically distressed than married men, yet the gap between the sexes in psychological distress is less pronounced for persons of Mexican identity, among whom the home and the family—women's traditional domain—are more highly valued. Our results, however, do not provide many clues to the mechanisms through which these cultural effects work, since a number of hypotheses concerning the differential effects of family and work roles on distress in the two cultures were not supported. First, we hypothesized that being a housewife who is not employed outside the home would increase psychological distress for Anglo women but not for Mexican women. This was not the case. Women who are employed are less psychologically distressed than housewives in both cultures. Second, we hypothesized that marital satisfaction would have a larger effect on psychological distress for Anglos than for Mexicans, since the Mexicans have a larger family network to fall back on for support if the marriage is not happy. This was not the case. Marital satisfaction has a large effect on psychological well-being for men and women in both cultures. Third, we hypothesized that the presence of children would increase distress for Anglo women but not for Mexican women. Unexpectedly, though, children had no detrimental effect on psychological well-being in these data, for either married men or women, whether Mexican or Anglo.[5]

There are two interesting findings concerning family and work roles that we think merit further investigation even though they have not explained the cultural variation in the relationship between sex and psychological distress in our study. In spite of the fact that significantly more Anglo wives than Mexican wives are employed, the Anglo women are no more likely to have a nontraditional division of labor in the home. Married women who work are still likely to have the major responsibility for cooking and cleaning. The Anglo family is less traditional than the Mexican family in one aspect—the wife's labor force participation—but no less traditional in terms of the division of labor in the home. Anglo women are not in traditional homes where the wife receives prestige for her role in the family, yet they are not in nontraditional homes where the husband and wife share the work (both outside and inside the home) equally.

The relative occupational prestige of the married men and women who are employed in the Mexican and Anglo groups also merits further investigation. Anglo men have the highest occupational prestige (529), with

[5] We hope future research reexamines these hypotheses in two cultural groups that are more diverse, since Mexican and Anglo culture—especially in the border cities of El Paso and Juarez—may be similar in many ways.

employed Anglo women being somewhat lower (502). However, among the Mexicans this trend is reversed: employed Mexican women have higher occupational prestige scores (362) than their male counterparts (274). These trends indicate that the relative status of Anglo men and women is different from the relative status of Mexican men and women. Among the Anglos, not only are the women accorded little prestige for their role in the family, but the men are accorded a great deal of prestige for their occupational roles. Among the Mexicans, however, the men have very low occupational prestige—lower even than the women who are employed—while the women receive more status for their position in the family.

The wife's participation in the labor force has complex effects on the family and on the couple's psychological well-being. When the wife works, both she and her husband are less satisfied with the marriage, possibly because the wife resents the fact that the husband does not help with the housework and because her work overload leaves less time for companionship. Reduced marital satisfaction increases psychological distress. However, the wife's employment also decreases both partners' psychological distress directly: for the woman, employment may indicate increased power, personal economic resources, and self-esteem, while for the man, a working wife may indicate an improved economic outlook, not so much by increasing income as by maintaining the accustomed standard of living.

American wives take care of the children, clean the house, and cook the meals, just as Mexican wives do. However, American wives find less support for the role of housewife and mother built into their way of life and find that a housewife and mother has much less prestige and authority than a husband and breadwinner. If a married woman gets a job to bolster the family income or to find self-expression through occupational achievement, or both, she finds that the wife is now more like a husband but the husband is not more like a wife. Tensions and contradictions result. In American society the interlocking expectations and obligations of husband and wife are held by tradition against the press of modern ways. In the future, tradition may yield, with husbands becoming more like wives by sharing in the housework and child-rearing duties, or tradition may hold and the tensions and contradictions persist.

REFERENCES

Baca Zinn, Maxine. 1975. "Political Familism: Toward Sex-Role Equality in Chicano Families." *Aztlán: Chicano Journal of the Social Sciences and the Arts* 6:13–26.
Brown, George W., and Tirril Harris. 1978. *Social Origins of Depression: A Study of Psychiatric Disorder in Women*. New York: Free Press.
Burke, Ronald J., and Tamara Weir. 1976. "Relationship of Wife's Employment Status to

Husband, Wife and Pair Satisfaction and Performance." *Journal of Marriage and the Family* 38:279–87.

Crowne, Douglas P., and David Marlowe. 1960. "A New Scale of Social Desirability Independent of Psychopathology." *Journal of Consulting Psychology* 24:349–54.

Elmendorf, Mary. 1977. "Mexico: The Many Worlds of Women." Pp. 127–72 in *Women: Roles and Status in Eight Countries,* edited by Janet Z. Geile and Audrey C. Smock. New York: Wiley.

Farris, Buford E., and Norval D. Glenn. 1976. "Fatalism and Familism among Anglos and Mexican Americans in San Antonio." *Sociology and Social Research* 60:393–402.

Gove, Walter R., and Michael R. Geerken. 1977. "The Effect of Children and Employment on the Mental Health of Married Men and Women." *Social Forces* 56:66–76.

Gove, Walter R., and Jeannette Tudor. 1973. "Adult Sex Roles and Mental Illness." *American Journal of Sociology* 78:812–35.

Keefe, Susan E., Amado M. Padilla, and Manuel Carlos. 1979. "The Mexican American Family as an Emotional Support System." *Human Organization* 38:144–52.

Langner, Thomas S. 1962. "A Twenty-two Item Screening Score of Psychiatric Symptoms Indicating Impairment." *Journal of Health and Human Behavior* 3:269–76.

Lomnitz, Larissa. 1977. *Networks and Marginality: Life in a Mexican Shantytown.* New York: Academic Press.

Madsen, William. 1964. *The Mexican Americans of South Texas.* New York: Holt, Rinehart & Winston.

Mirandé, Alfredo. 1977. "The Chicano Family: A Reanalysis of Conflicting Views." *Journal of Marriage and the Family* 39:747–56.

Mirowsky, John, and Catherine E. Ross. 1983. "Paranoia and the Structure of Powerlessness." *American Sociological Review* 48:228–39.

Murillo, Nathan. 1976. "The Mexican American Family." Pp. 15–25 in *Chicanos: Social and Psychological Perspectives,* edited by Carroll A. Hernandez, Marsha J. Haug, and Nathaniel Wagner. St. Louis: Mosby.

Pearlin, Leonard. 1975. "Sex Roles and Depression." Pp. 191–208 in *Life Span Developmental Psychology: Normative Life Crises,* edited by Nancy Datan and Leon H. Ginsberg. New York: Academic Press.

Radloff, Lenore. 1975. "Sex Differences in Depression: The Effects of Occupation and Marital Status." *Sex Roles* 1:249–65.

Renne, Karen S. 1970. "Correlates of Dissatisfaction in Marriage." *Journal of Marriage and the Family* 32:54–66.

Schafer, Robert B., and Patricia M. Keith. 1980. "Equity and Depression among Married Couples." *Social Psychological Quarterly* 43:430–35.

Vanfossen, Beth. 1981. "Sex Differences in the Mental Health Effects of Spouse Support and Equity." *Journal of Health and Social Behavior* 22:130–43.

Weissman, Myrna M., and Eugene S. Paykel. 1974. *The Depressed Woman: A Study of Social Relationships.* Chicago: University of Chicago Press.

Wheaton, Blair. 1982. "Uses and Abuses of the Langner Index." Pp. 25–54 in *Psychosocial Epidemiology: Symptoms, Illness-Behavior, and Help-Seeking,* edited by David Mechanic. New York: Neale Watson.

Commentary and Debate

To conserve space for the publication of original contributions to scholarship, the comments in this section must be limited to brief critiques. They are expected to address specific errors or flaws in articles and reviews published in the *AJS*. Comments on articles are not to exceed 1,500 words, those on reviews 750 words. Longer or less narrowly focused critiques should be submitted as articles. Authors of articles and reviews are invited to reply to comments, keeping their replies to the length of the specific comment. The *AJS* does not publish commenters' rebuttals to authors' replies. We reserve the right to reject inappropriate or excessively minor comments.

RETURN TO INDIANAPOLIS: A COMMENT ON THE LOGIC OF MODEL CONSTRUCTION[1]

Baron ("Indianapolis and Beyond: A Structural Model of Occupational Mobility across Generations," *AJS* 85 [January 1980]: 815–39) applies a structural model for Rogoff's Marion County 1910–40 data to a mobility matrix from the 1973 Occupational Changes in a Generation (OCG) sample. If such a model (Hauser 1978) forms an appropriate baseline, on what grounds are we to accept the particular model which is to structure the comparison? The question is raised most readily by presenting alternative structural models of disparate import and asking how Baron could choose among them. The point could equally be made by noting that the particular model is not derivable from the data and theoretical assumptions.

Baron presents a nine-level[2] structural model with three asymmetries for Indianapolis:[3] "A model was adopted when major modifications clearly threatened substantive validity, parsimony, or goodness of fit, and where

[2] It is not clear why we focus on a nine-level model; theory says nothing about nine. The "elbows" in the graph of log G^2 by levels for located "best-fitting" symmetrical models lie at 4 and 7; since no published minimization algorithm exists, this, also, is inconclusive.

[3] Baron's fig. 1 (p. 820) is Baron's (1977) fig. 3a (p. 74) with (4,7) and (7,4) moved from level 9 to level 8 (slightly reducing "fit" to Indianapolis and increasing OCG fit).

minor revisions did not appreciably affect the portrait of mobility structure obtained" (p. 820). Cautioning against the temptation to "overfit," he notes, "The resultant model (not, incidentally, the best-fitting one) was accepted because of its interpretability, parsimony, and close fit" (p. 820). Baron (1977) is no more explicit on the logic of the selection. The investigator seeking to replicate will not necessarily reach the same place, nor will that investigator and Baron be able to articulate the grounds of their differences.

Believing that the mobility process is symmetrical, Baron first fits the data smoothed by both time and symmetry;[4] he then fits the data smoothed by time only. Accepting that assumption, consider the upstart symmetric nine-level model in (my) table 1 (which differs substantially[5] from Baron's model, particularly in the populous category 5). Baron's symmetric model (from which Baron's fig. 1 [p. 820] is derived) has a G^2 of 39.4 applied to the data smoothed by time and symmetry, and 97.7 applied to the data smoothed by time only, while the upstart achieves values of 14.1 and 72.4. By introducing three asymmetries, Baron reduces the G^2 for the time-smoothed data to 75.7; by introducing three asymmetries,[6] I obtain 51.5. My symmetric model is a better fit, and the asymmetric model a markedly better fit, than Baron's asymmetric model. Why, then, does Baron choose to discuss his particular model?

I share the view that goodness of fit should not override theoretical concerns, but it is not clear how these would allow one to choose. Both models lend themselves to substantive interpretations.[7] Table 1 (in contrast to Baron's) sees the density of self-recruitment in all occupations as high. Again in contrast, it explicably sees outflow from protective service as restricted, and reveals relatively high, and proper, interchange between unskilled labor and farming. Both score well on parsimony and fit (table 1 is the tighter fit and, being symmetrical, more parsimonious). But both are equally vacuous; neither is tethered by assumptions, procedures, and data.

The models do behave differently when the 1973 OCG data are used. Then Baron's asymmetrical model achieves a G^2 of 636.6 (its symmetrical source yields 652.6). In my models the situation is reversed; the

---

[4] Those not sharing Baron's conviction that mobility processes are symmetric will be uneasy at smoothing data prior to model fitting, but the move is logically proper.

[5] Any index of similarity of parameters between models should weight for cell frequency.

[6] Cell (2,4) moves from level 3 to 4; (5,8) from 7 to 6; (1,5) from 6 to 5; and (5,1) to 7.

[7] Baron's geographical metaphor of "provinces separated by several 'gullies' indicating blockages to intergenerational movement" (p. 821) is misleading. If the density of movement is indicated by the appropriate level, the densities of intervening cells are *not* additional measures of additional blockages.

symmetrical model has a G^2 of 644.9, the asymmetrical 701.2; this would say that the Indianapolis asymmetry is not present in the OCG data, a finding running counter to Baron's.[8] If temporal stability is treated as an assumption, some minor[9] modifications of table 1 yield a model which, on fit to the Indianapolis time-smoothed data, is equivalent to Baron's (77.9 compared with Baron's 77.8), and on the OCG data is a tighter fit (627.1 compared with 636.6), with the further advantage of only two asymmetries.

Other models could be adduced.[10] The point is that Baron has presented one particular solution without indication that there exist, within the formal and theoretical constraints adopted, competing models yielding substantively different conclusions on mobility. These twin characteristics (of unique report and multiple contenders)—which combined are dangerous—characterize all current published uses of such a structural model. The general problem for Baron is, Can he provide a set of rules or assumptions which should lead us to attend to the particular solution? Consider a regression model: the parameters in the equation are

TABLE 1

NINE-LEVEL SYMMETRIC MODEL

	SONS									
FATHERS	1	2	3	4	5	6	7	8	9	10
1. Professional	1									
2. Semiprofessional	3	1								
3. Proprietors, managers, and officials	3	4	2							
4. Clerical and sales	3	3	3	2						
5. Skilled labor	6	6	6	4	3					
6. Semiskilled labor	8	7	6	5	4	3				
7. Unskilled labor	9	9	8	7	5	4	2			
8. Protective service	9	9	6	6	7	6	6	3		
9. Personal service	8	4	5	5	5	4	5	9	2	
10. Farming	7	9	6	7	6	6	4	6	5	1

[8] G^2 values applying Indianapolis models to OCG data are all large; Baron's asymmetric model (p. 820) can be outperformed by an OCG specific three-level symmetric one. This weakens Baron's claim to have provided "extremely persuasive evidence of the relevance of the Indianapolis mobility experience to contemporary trends" (p. 826).

[9] Cells (5,2), (2,5), and (10,5) move from level 6 to 5, and (6,4) from 5 to 4.

[10] If fit to the time-and-symmetry-smoothed Indianapolis data is taken as preeminent, it is possible to locate a symmetric nine-level model which achieves a G^2 of 8.8 compared with Baron's 39.4 (and fits the time-smoothed data better); this fits the OCG data badly and thus implies shift in structure between time points.

interpretable only insofar as there exist rules which generate them; without the rules they reduce to mere numbers, uninterpretable digits. Presenting an interpretable result which is "not the best-fitting equation" can only be done by adducing the grounds on which some specified constraint has been imposed. Lack of specification, I suggest, characterizes Baron's account.

There are two contenders for the restoration of logical integrity. One method would be to attend to the description of row and column effects[11] as measuring "supply" and "demand" (p. 828) and to require plausible estimates. This would necessitate some criterion, external to the pattern of levels chosen, for assessing plausibility. Note that, assuming origin marginals are independent of the structure of the table, a counterfactual world cannot be constructed in which the pattern of demand has moved while supply and density have remained constant[12]; this mathematical interconnection within the model fits badly with the suggested interpretation that invokes analytically distinct terms. Further, there is nothing in the structure of the model itself to legitimate translation as supply and demand, and, a fortiori, despite my earlier statements, no grounds for reading the levels parameters as mobility net of such effects. Another method would be to expand the formal component in the specification, setting constraints on the arrangement of levels and possibly on the relative magnitudes of parameters. Such an approach is found in Hope (1981) (and in a milder form in Duncan [1979]), but, whatever its merits, it would not retain the essence of the structural model (Hauser 1981) in which the structure is to be located by the conjunction of minimal formal constraints and the data and is the "outcome of a search procedure" (Hauser 1979, p. 420). The moves to logical integrity save the model by subverting it.[13] Pöntinen (1982, p. 105), on grounds akin to those in Macdonald (1981, p. 562), has argued that such models "can only weakly support the ideas which they reflect"; both interpret negatively a set of exact equivalences which Goodman (1979, p. 805) reads positively. The present argument rests on nonformal near-equivalences (not peculiar to the table under discussion) but also leads to pessimism. In conclusion, it should be stressed that no charge of individual malfeasance is laid against Baron whose substantive

[11] Baron's usage of the ratio of row and column effects (p. 828) is misguided, for these depend on the norming chosen (Holt 1979).

[12] With the model in additive (logarithmic) form, if origin (row) marginals and densities (interactions) are taken as given, and cell entries otherwise allowed to vary, an increase of x in the demand parameter for some specified column j entails row-specific modifications in supply by $\ln(n_i) - \ln[n_{i.} + (e^x - 1)n_{ij}]$. While the empirical interconnection of demand, supply, and density is of obvious interest, this particular mathematical entailment within the model requires explication.

[13] The argument is developed in Macdonald (1983b) and its general import discussed in Macdonald (1983a).

Commentary and Debate

exposition is informed and informative; the gap in his argument is a gap in the logic of the structural model.

K. I. MACDONALD

Department of Social and Administrative Studies
and *Nuffield College, Oxford*

REFERENCES

Baron, J. N. 1977. "The Structure of Intergenerational Occupational Mobility: Another Look at the Indianapolis Mobility Data." Center for Demography and Ecology Working Paper 77-35. Madison: University of Wisconsin.
Duncan, O. D. 1979. "How Destination Depends on Origin in the Occupational Mobility Table." *American Journal of Sociology* 84:793–803.
Goodman, L. A. 1979. "Multiplicative Models for the Analysis of Occupational Mobility Tables and Other Kinds of Cross-Classification Tables." *American Journal of Sociology* 84:804–19.
Hauser, R. M. 1978. "A Structural Model of the Mobility Table." *Social Forces* 56:919–53.
———. 1979. "Some Exploratory Methods for Modelling Mobility Tables and Other Cross-classified Data." Pp. 413–58 in *Sociological Methodology 1980,* edited by K. F. Schuessler. San Francisco: Jossey-Bass.
———. 1981. "Hope for the Mobility Ratio." *Social Forces* 60:572–84.
Holt, D. 1979. "Log-Linear Models for Contingency Table Analysis: On the Interpretation of Parameters." *Sociological Methods and Research* 7:330–36.
Hope, K. 1981. "Vertical Mobility in Britain: A Structured Analysis." *Sociology* 15:19–55.
Macdonald, K. I. 1981. "On the Formulation of a Structural Model of the Mobility Table." *Social Forces* 60:557–71.
———. 1983a. "Exploratory Data Analysis: A Process and a Problem." Pp. 256–84 in *Data Analysis and the Social Sciences,* edited by D. McKay, N. Schofield, and P. Whiteley. New York: St. Martin's Press.
———. 1983b. "On the Interpretation of a Structural Model of the Mobility Table." *Quality and Quantity* 17:203–24.
Pöntinen, S. 1982. "Models and Social Mobility Research: A Comparison of Some Log-Linear Models for a Social Mobility Matrix." *Quality and Quantity* 16:91–107.

REPLY TO MACDONALD[1]

> Each man is given a scientific heritage plus a continuing barrage of sensory stimulation; and the considerations which guide him in warping his scientific heritage to fit his continuing sensory promptings are, where rational, pragmatic. [QUINE 1961, p. 46]

My article sought to demonstrate three things. First, a model disentangling changes in occupational structure from tendencies toward occupa-

687

tional (im)mobility inspires different conclusions from those Rogoff (1953) reached in her influential work. Second, the distinctiveness of mobility regimes according to race, nativity, age, and time seems attributable to different patterns of occupational origins and destinations, while the "sociometry" of occupational roles appears invariant. Third, my account has some generalizability and demonstrates the utility of *cross-validation* in mobility studies.

While mobility analysts have tended more toward induction than deduction, their models—and mine—are not necessarily "equally vacuous." It might seem obvious, for instance, that professionals and farmers should recruit disproportionately from within their ranks. However, this recognition requires some familiarity with historical mobility trends and an appreciation of the economic and nonmonetary factors encouraging occupational inheritance within those situses.

In selecting a model, I used backward selection procedures identical to model search strategies employed routinely by data analysts. Some of my assumptions were substantiated and others, such as the hypothesis of high inheritance in all occupations, were invalidated and required modifications. The model I eventually retained partitions the mobility table into regions of (almost) uniformly high, medium, or low density. This seemed desirable because it yielded intriguing insights—for example, regarding the relationship between mental and manual work roles—and facilitated the model's interpretation.[2]

To see whether my assumptions were idiosyncratic, cross-validation was imperative. As Mosteller and Tukey note, "When fresh [data] gathering is not feasible, good results can come from going to a body of data that has been kept in a locked safe where it has rested untouched and unscanned during all the choices and optimizations" (1977, p. 38; also see Duncan 1982, p. 961). Accordingly, the model was subjected to the strictest test of external validity possible given differences in sampling, geography, and temporal change.[3]

Of course, others might produce plausible alternative models (e.g., Featherman and Hauser 1978; Breiger 1981). My cross-validation simply suggested that certain important aspects of the mobility process transcend

[2] Macdonald takes issue with my geographical metaphor, apparently misreading me (see my p. 821). The geographical imagery does take the ordering of categories as given, since otherwise the "coordinates" of the territory would be arbitrary. Given that ordering, my fig. 1 shows where there is a lot of net (im)mobility. The "gullies" do not refer to rows and columns but, rather, to clusters of row-column combinations in which there is considerably less exchange than would be expected given the origin and destination effects.

[3] Macdonald questions my "conviction" that mobility processes are symmetric. The assumption of symmetric interactions was based not on conviction but on goodness-of-fit tests confirming the hypothesis of symmetry in Rogoff's data (see p. 820; Baron 1977, p. 42).

time and space. A more "tailor-made" account for each mobility matrix would yield different results, and Macdonald's examples document that fact. While Macdonald demonstrates that other models exist (this is always true), he offers no reason to impugn my conclusions.

. Furthermore, notwithstanding some differences, Macdonald's results are remarkably similar to mine. If the row (origin) parameters from his model are compared with mine, the correlation coefficient equals .99; the same is true of the column (destination) parameters. Even our level effects correlate .84 across cells.[4] In short, overall our models tell very similar stories about the relative pushes, pulls, and proximities among occupational categories, although specific parameter estimates may differ.

Macdonald quotes from my unpublished master's thesis (Baron 1977) but overlooks chapter 4, which shows that row and column effects under my model depict temporal changes in occupational distribution that are overwhelmingly consistent with the history of Indianapolis during the period spanned by Rogoff's data.[5] That exercise sought to do exactly what Macdonald accuses me of not doing, namely, demonstrate how my exploratory results correspond to the real world.

My primary concern is that a model tell a sensible and powerful story. On this score I did not find Macdonald helpful. For instance, he claims that his model, in contrast with mine, reveals "high, and *proper*, interchange between unskilled labor and farming" (emphasis added). Proper in what sense? If Macdonald has a reason for positing higher rates of movement among these situses, that would be a useful contribution. In that regard, I am encouraged by recent generalizations and extensions of the methods I employed (Breiger 1981; Clogg 1981; Hout 1984; Yamaguchi 1983).

My article was a point of origin, not a destination. Macdonald's model, for example, reveals a more consistent gradation among skilled, semiskilled, and unskilled labor. This suggests that mobility flows may reflect a skill or status hierarchy more than my analyses suggested. He could reformulate a model consistent with that proposition and perhaps produce

---

[4] These parameter comparisons are between Macdonald's table 1 and my nine-level symmetric model fitted to the Indianapolis data smoothed by time and symmetry (see Baron 1977, pp. 48–69). Weighting by the relevant cell counts only reduces the correlations to .96 (rows), .98 (columns), and .80 (levels).

[5] Macdonald objects to my using ratios between row and column effects in discussing intergenerational shifts because they depend on the specific normalizations adopted. All comparisons depend on some normalization. By deviating row and column effects from the grand mean, I opted to compare the intergenerational shift in a particular category with the average shift among all categories. This seemed more reasonable than performing comparisons against, say, some arbitrarily omitted occupational category. In retrospect, given the additive metric of the row and column effects, it would have been more consistent to compute row-column differences, though this would not have changed the story.

American Journal of Sociology

a more powerful and interpretable account than I did (see, e.g., Hout 1984; Yamaguchi 1983). I would applaud the effort. Therefore, I do not agree that the class of model I used precludes subsequent comparisons, refinements, or invalidation, and I eagerly await a positive contribution from Macdonald.

JAMES N. BARON

Stanford University

REFERENCES

Baron, James N. 1977. "The Structure of Intergenerational Occupational Mobility: Another Look at the Indianapolis Mobility Data." Center for Demography and Ecology Working Paper 77-35. Madison: University of Wisconsin.
Breiger, Ronald L. 1981. "The Social Class Structure of Occupational Mobility." *American Journal of Sociology* 87 (November): 587–611.
Clogg, Clifford C. 1981. "Latent Structure Models of Mobility." *American Journal of Sociology* 86 (January): 836–68.
Duncan, Otis D. 1982. "Review Essay: Statistical Methods for Categorical Data." *American Journal of Sociology* 87 (January): 957–64.
Featherman, David L., and Robert M. Hauser. 1978. *Opportunity and Change*. New York: Academic Press.
Hout, Michael. 1984. "Status, Autonomy, and Training in Occupational Mobility." *American Journal of Sociology*, vol. 89 (May).
Mosteller, Frederick, and John W. Tukey. 1977. *Data Analysis and Regresion*. Reading, Mass.: Addison-Wesley.
Quine, Willard V. O. 1961. *From a Logical Point of View*. 2d ed. Cambridge, Mass.: Harvard University Press.
Rogoff, Natalie. 1953. *Recent Trends in Occupational Mobility*. New York: Free Press.
Yamaguchi, Kazuo. 1983. "The Structure of Intergenerational Occupational Mobility: Generality and Specificity in Resources, Channels, and Barriers." *American Journal of Sociology* 88 (January): 718–45.

COMMENT ON BORNSCHIER, CHASE-DUNN, AND RUBINSON

Over the past few years Volker Bornschier has written or coauthored a number of articles on the effect of foreign investment on rates of economic growth in the less developed countries. The most influential of these (Bornschier, Chase-Dunn, and Rubinson's "Cross-national Evidence of the Effects of Foreign Investment and Aid on Economic Growth and Inequality"), which attempted to sum up and transcend the existing studies on the topic, appeared in the *American Journal of Sociology* (*AJS* 84 [November 1978]: 651–83). Bornschier's central thesis is that the *stock* (but not the flow) of foreign investment in a country has a long-term

690

negative effect on its rate of growth, a thesis that goes against both what a glance at comparative statistics seems to show and what the leaders of nearly all less developed market economies believe. Bornschier's most recent work takes his analysis to the reductio ad absurdum conclusion that stock of foreign investment in both manufacturing and mining is negatively correlated with rates of growth, while there is no significant relation between investment in petroleum and the rate of growth (Bornschier 1981).

A comparison of the 50 major less developed market economies over the 1970–77 period shows that the 25 with the lowest ratios of foreign investment stock to GDP had a rate of growth of 2.7% per capita, while the 25 with the highest concentrations had a rate of growth of 3.4% (25% greater) (see Szymanski 1981, p. 330).

Bornschier's argument is based on the assumption that such results are spurious and produced by the "extraneous" relation among such factors as the level of domestic capital investment, rates of growth, and the level of exports, on the one hand, and the parallel association of stock of foreign capital with the level of domestic capital investment and the level of exports, on the other. Bornschier claims that the real effect of foreign investment stock on growth can be revealed only when the effect of these other factors is taken out or controlled for. When he does this (and controls for still other key factors that allegedly produce a spurious association between the stock of foreign investment and rates of growth such as the rate of growth in foreign investment) he finds, not surprisingly, that there is a slight negative relation between foreign investment stock and the rate of growth.

Bornschier's work is based largely on a very serious and fundamental fallacy. When one controls for a variable, one can be doing one of two things: taking out the spurious effect of an extraneous factor in order to determine the real relationship between A and B, or proving that the effect of A on B occurs through the (intermediate) variable controlled for. Those who maintain that there is a positive relation between foreign investment stock and rate of growth generally argue that foreign investment affects economic growth by, for example, expanding the level of hard currency exports that can then be used to import high technology, mobilizing private and public funds in partnership with local capital, and generating profits that are then reinvested in the local economy. Thus, it is not surprising that Bornschier, by controlling for these factors, finds a negative association between the stock of foreign investment and the rates of growth. What Bornschier really shows is not that there is a negative association between the stock of foreign investment and the rate of growth but, rather, that the primary paths through which the positive effect of foreign investment is exercised are through foreign investment's direct

691

impact on expanding export earnings and mobilizing local capital (including reinvested earnings) (World Bank 1979, tables 2 and 5).

An equally serious fallacy in the Bornschier argument is his use of the rate of growth in foreign investment as another control variable. Foreign capital is very likely to flow rapidly into a country where existing investments are the most profitable. This is highly likely to be the case in economies that are growing rapidly, that is, there is a high positive correlation between the rate of growth in foreign investment and the overall rate of growth (in foreign plus domestically owned investments.) Thus, controlling for the rate of growth in foreign investments means that most of the effect of foreign investment on economic growth is taken out. Even if the positive effect of foreign investment on economic growth that is exercised through expanding export earnings and mobilizing domestic capital were not taken out, it would not be surprising that Bornschier could still "prove" by such fallacious procedures that the stock of foreign investment had a negative effect on growth.

Bornschier et al.'s 1978 article commits a number of other errors which also bias his model in the direction of necessarily finding a negative relation between rates of growth and foreign investment stock. For example, his measure of the stock of foreign capital (FORCAP) as essentially the composite of capital stock divided by energy consumption and capital stock per capita is highly biased in favor of his hypothesis. The first measure takes out part of the positive effect on growth, since foreign investment is disproportionately invested in sectors which disproportionately produce or consume energy (or have close linkages, financial and otherwise, with it). Furthermore, the second measure confounds poor countries where a large percentage of the GDP is foreign owned, with rich countries where only a small percentage of the economy is foreign owned.

It is important to note that over the 1966–76 period, 49% of all net new investment funds of U.S. transnational corporations in the less developed countries were reinvested earnings, 50% were funds acquired locally, and only 1% were funds newly transferred from the United States (see U.S. Department of Commerce 1979, tables 12 and 13). It is clear, then, that foreign capital (or at least U.S. capital, which represents roughly 50% of all foreign capital in the LDCs) exerts its influence mostly through reinvested earnings and mobilization of local capital rather than through any new infusion of funds from the United States. Thus, controlling for foreign capital's impact on domestic capital formation ignores a fundamental process by which growth is affected.

Likewise it should be noted that the most rapidly growing LDC market economies tend to have the highest ratios of exports to GDP (this is true for both the rapidly growing manufacturing export economies and the OPEC petroleum exporting countries). For example, the ratio of exports

of goods and nonfactor services to GDP in 1977 for Hong Kong was 98%, for Taiwan 54%, for South Korea 40%, and for Singapore 160% (compared with 20% for all the "middle-income" countries and 14% for all "low-income" countries). These countries in 1970–77 had rates of growth in industry of 6.8%, 12.2%, 17.0%, and 8.6%. Likewise, Saudi Arabia and Iran had a ratio of exports of goods and nonfactor services to GDP of 78% and 34%, while they had rates of growth in GDP of 12.9% and 7.8%, respectively (compared with the overall average of 6.1%).

Although Bornschier's analysis is clearly fallacious, his conclusions about the long-term effect of foreign capital stock are not altogether incorrect. It is true that primary dependence on foreign investment to promote rapid economic growth slows down economic growth in relation to what could be produced by other models of development. Within the parameters of primary reliance on markets, private capital, and foreign investments, there are forces which inhibit economic growth (not the least of which is monopolization). Within the set of such countries, those with more foreign investment grow faster than those with less, since the potentially dynamic factors of central planning, state funded investment, and local private industrial capital are inhibited by integration into the world market. However, those countries which rely primarily on local state-funded accumulation and central planning as well as (to a lesser extent) those technically advanced and integrated economies which have a dynamic domestic private sector grow more rapidly than those less developed countries without high concentrations of foreign capital, central planning, or a high-technology domestic private sector.

Since World War II the centrally planned economies of Eastern Europe have had the most rapidly increasing rates of growth of any region in the world. For example, in the 1960–70 period their average rate of growth in GDP per capita was 5.4% a year, compared with 2.6% for the less developed economies and 3.8% for the developed market economies. In the 1970–77 period it averaged 5.1% a year, compared with 3.2% for the less developed countries and 2.5% for the developed market economies (see United Nations 1979, table 6.2). It might be noted that over the entire 1960–77 period, the country with the most rapid rate of growth in GNP per capita in the entire world was Rumania, which grew at the annual average rate of 8.5% (this compares with Japan's 7.7% and Hong Kong's 6.5%) (see World Bank 1979, table 1).

It should be noted that the other set of countries with very rapid rates of growth in the post-1960 period has been the Asian economies with dynamic, primarily locally owned manufacturing sectors (in addition to the two just named, Singapore 7.5%, Taiwan 6.2%, and South Korea 7.4%). But in the absence of either central planning, as in Eastern Europe, or a dynamic "national bourgeoisie," as in the Asian manufacturing export economies, concentrations of foreign investment produce higher rates of

American Journal of Sociology

economic growth than their absence, primarily because of foreign investment's effect on expanding hard currency–earning exports and mobilizing local capital.

ALBERT SZYMANSKI

University of Oregon

REFERENCES

Bornschier, Volker. 1981. "Dependent Industrialization in the World Economy." *Journal of Conflict Resolution* 25 (September): 371–400.
Szymanski, Albert. 1981. *The Logic of Imperialism*. New York: Praeger.
U.S. Department of Commerce. 1979. *Sources and Uses of Funds of Majority Owned Foreign Affiliates of U.S. Companies 1966–1976*. Bureau of Economic Analysis Staff Paper. Washington, D.C.: Government Printing Office.
United Nations. 1979. *Trade and Development Statistics*. New York: United Nations.
World Bank. 1979. *World Development Report*. New York: World Bank.

REPLY TO SZYMANSKI

Albert Szymanski disputes our conclusions concerning the role of multinational corporations (MNCs) in national economic growth (in addition to the *AJS* article commented on, see Bornschier, Chase-Dunn, and Rubinson 1980; Bornschier 1981). We strongly disagree with him when he maintains that our research "is clearly fallacious." Our reply focuses on his two main contentions: (1) that there is a positive association between dependence on MNC investment and economic growth and (2) that our results are due to controlling out the positive effects of MNC investment. For brevity we will refrain from discussing our other main finding—that MNC dependence is associated with greater income inequality—and from commenting extensively on the ironies of intellectual history which find a Marxist sociologist (Szymanski) the main critic of research which demonstrates that multinational corporations distort and retard economic development in peripheral countries.

Our main independent variable is MNC penetration, or the proportion of all capital stock in a country which is foreign owned and controlled. Szymanski contends that our conclusion that MNC penetration has long-term negative effects on economic growth goes against "what a glance at comparative statistics seems to show." By this he means that in a sample of 50 peripheral countries dichotomized on his indicator of MNC penetration, the group with greater dependence on foreign capital has a higher

694

mean percentage of economic growth than the less dependent half of the sample.

This is the single piece of new evidence contained in Szymanski's critique. We do not know how he obtained this result. Our own foray into the comparison of subsample means (reported below) reveals very different results. We do not share Szymanski's predilection for methodologically simplistic analyses but adopt his methods for purposes of comparison.

We must make a distinction which Szymanski apparently does not understand. We want to distinguish between the overall dependence of a national economy on foreign capital holdings (PEN) and the fresh flows of foreign investment capital which arrive in a particular time period (FDI). Thus our measure of dependence on foreign corporations (PEN) is the total book value of foreign-held capital stocks as estimated by the OECD in 1967 weighted by size measures of the penetrated country. Our measure of new investment inflows (FDI) is the difference between the value of foreign-held stocks in 1972 from that value in 1967.[1] This distinction is important because all theories agree that the immediate impact of new investments is to increase GNP as land, labor, and other commodities are purchased and put into operation. What is in dispute is the long-run consequence of national dependence on foreign investment. Thus we want to study this dependence net of the immediate growth spurt brought about by new inflows of foreign capital.

In a sample of 88 peripheral and semiperipheral countries,[2] the average compound rate of GNP per capita growth between the end of 1965 and 1977 was 2.6% (standard deviation = 2.3). A look at the overperforming countries—here defined as those which had a growth rate of 4.5% and more—reveals the following.

In that period 17 countries had a growth rate of 4.5% or more.[3] The greatest growth occurred in South Korea (8.7%). For the group of overperforming countries the average figure for PEN in 1967 was 38.5 as

[1] The different weights for PEN and for FDI are described by Bornschier (1981, annex). Outliers on both variables have been recoded. For a discussion of alternative procedures to tackle the problem of outliers see Bornschier et al. (1980, p. 191, n. 26) and Bornschier (1982).

[2] This refers to the subsample of 88 peripheral and semiperipheral countries which was analyzed together with a world sample by Bornschier (1981). The split that separates the subsample from the full sample is based on the level of GNP per capita in the post–World War II period up to 1965 (for the reasons see Bornschier [1980b]). With a few exceptions (Japan, Italy) the countries of the subsample are the ones conventionally regarded as "less developed or developing" countries.

[3] The following countries have been coded as overperforming (in the order of their growth rates): South Korea, Rumania, Singapore, Saudi Arabia, Iran, Hong Kong, Japan, Taiwan, Brazil, Yugoslavia, Greece, Portugal, Tunisia, Dominican Republic, Turkey, Thailand, Indonesia.

compared with 49.0 (standard deviation = 35.1, range 0–105) for the whole group of 88 peripheral and semiperipheral countries. The over-performers had an average figure of FDI between the end of 1967 and 1973 of 87.8 as compared with 65.8 for the whole group. This demonstrates that the level of dependence on foreign investment in fast-growing countries was lower and the inflow of new foreign capital was greater than in the whole group of peripheral and semiperipheral countries.[4] This contradicts Szymanski's claim that a simple inspection shows that dependence and growth are positively related.

The comparison above neglects the important fact that PEN is correlated with higher initial levels of GNP per capita and that this is also true of overperforming countries. The correlation between PEN in 1967 and logged GNP per capita in 1965 is .42, and the correlation between logged GNP per capita and the subsequent economic growth rate is .41.

Of the 17 overperforming cases, 14 had an initial GNP per capita above the mean of the whole subsample.[5] For the more developed of the sample of 88 peripheral and semiperipheral countries (those above the mean on GNP per capita in 1965), the average PEN is 59.8, which is much higher than the average score for the whole subsample of 88 (49.0). Thus, while the richer countries in the subsample have, on the average, a higher level of penetration by foreign capital (59.8), the average for the group of overperforming countries (which are mostly in this more developed group) is substantially lower (38.5).[6]

Therefore no evidence supports Szymanski's view that countries which have greater dependence on foreign investment also have higher growth rates. Quite the opposite is true. There is, however, a correlation between FDI and high growth.[7] This, however, is not in dispute.

The simple demonstration above also holds for foreign capital dependence specifically in the manufacturing sector. This contradicts what Szymanski and many other theorists have argued about the developmental effects of the "new dependency" (see Bornschier et al. 1980, p. 185; Bornschier 1981, p. 390).

[4] The zero-order correlations reveal these same relationships in the sample of 88 countries. The zero-order correlation between PEN in 1967 and growth 1965–77 is −.04 and that between FDI 1967–72 and growth 1965–77 is .19, while the intercorrelation between PEN and FDI is .40.

[5] For 1965 the mean of the logged GNP per capita figures for the 88 countries is 2.6; GNP per capita is in constant market prices and U.S. dollars of the base period 1975–77 (World Bank figures). The three overperforming cases which had an initial income level below the mean are South Korea, Thailand, and Indonesia.

[6] Since FDI is not substantially related to initial GNP per capita in the 1967–73 period, the comparison of average figures for FDI for overperforming countries has not been corrected for the level of GNP per capita.

[7] The correlation between PEN and FDI for the subsample of 88 is .40 (the same as in the full sample).

Thus, our conclusion is that rapidly growing peripheral and semiperipheral countries during the mid-1960s and mid-1970s were the ones with higher initial GNP per capita, with lower initial stocks of foreign investment, and with more foreign investment inflows during the period. This is consistent with our theoretical reasoning and with our previous, more sophisticated, empirical analyses.

Szymanski's second main criticism refers to what he calls a "fundamental fallacy." He maintains that in controlling for GDI (Gross Domestic Investment) and Exports (each of them as a percentage share of total GDP), we have taken out the positive effects MNCs should have on growth. In terms of our path analysis, he claims we have removed the positive *indirect* effects. Although Szymanski presents no evidence for his argument, logically his point could be right, but theoretically and empirically he is wrong. In cross-national comparison there are no indirect positive effects associated with these variables since PEN does not increase GDI or Exports.

In the first place we can show that the effect of PEN is significantly negative when one does not control for Exports and GDI as shares of GDP, although the negative effect is then somewhat smaller.[8] But this does not imply that there are indirect positive effects of PEN on growth via Exports and GDI. Obviously MNCs have in the past invested more capital in countries which have higher shares of Exports and GDI and which are richer with regard to GNP per capita. And high initial GNP per capita, Exports, and GDI may be—net of MNC presence—growth-promoting factors.

In other words, we maintain that our controls take out only the spurious positive association of PEN with growth due to its correlation with higher levels of Exports, GDI, and initial GNP per capita. This is easy to show since dependence on MNCs does not add to Exports, to GDI, or to GNP

[8] For example, stepwise regression in the sample of 88 on two levels (first level excluding Exports and GDI as shares of GDP) shows the following results (for the variables see Bornschier [1981]). First the size of the modern market enters (which is completely unrelated to PEN), then FDI follows, and after that PEN (with a zero-order correlation of .40 with FDI). The negative effect of PEN becomes more substantial when the logged initial level of GNP per capita enters the equation (the zero-order correlation with PEN is .42). A curvilinear specification of the GNP per capita effect adds to the predictive power of PEN. At the end of this first level of stepwise regression, PEN is clearly statistically significant with a β of − .29 and a *t*-ratio of 2.54. At the second level Exports as a share of GDP enter first with a significant contribution to growth (however, the zero-order correlation with PEN is high, .62, indicating that only exports which do not go together with a dominant role of MNCs add to growth). Controlling for Exports adds substantially to the predictive power of PEN. Gross Domestic Investment enters last (with a positive, although not statistically significant effect), and it increases only marginally the negative weight of PEN (zero-order correlation with PEN is .26). In the final equation PEN is by far the most significant single predictor of growth for the 88 peripheral and semiperipheral countries with a β of − .48 and a *t*-ratio of 3.95.

American Journal of Sociology

per capita growth. Panel analyses with Exports and GDI as dependent variables reveal no significant effects of PEN. Actually the signs of the coefficients are negative in both equations but not statistically significant.

The same conclusion is reached when growth rates instead of panel analyses are used. The effect of PEN in 1967 on the yearly growth rate of GDI from 1965 to 1973 is negative but not significant,[9] whereas FDI 1967–73 is significantly related to total growth of investment.

The effect of PEN 1967 on Export growth between 1970 and 1975 is not substantial, but PEN has a negative effect, whereas FDI has a significant positive effect.[10] These results hold for the world sample of countries as well as for the subsample of 88 peripheral and semiperipheral ones, and also for Export growth between 1965 and 1970 (which was, on the whole, much smaller than in the 1970–75 period).

Since dependence on MNCs does not cause increases in Exports and GDI and since these may have positive growth effects,[11] it is necessary to control these variables in order to obtain an unbiased estimate of the effects of PEN on subsequent growth. Thus it must be concluded that cross-national evidence does not support Szymanski's critique.

Dependence on foreign investment by multinational corporations does indeed have long-run retardant effects on national economic growth. It is understandable that scholars close to the ideologies proffered by the great firms to justify their operations might doubt this evidence. Of more interest is the fact that many Marxists and dependency theorists who are concerned to understand uneven development and exploitation have ignored or repudiated the evidence above. Unfortunately, when strong interests are involved, evidence has little effect on attitudes, and this is true among social scientists as well as other mortals.

VOLKER BORNSCHIER

University of Zurich

CHRISTOPHER CHASE-DUNN

Johns Hopkins University

REFERENCES

Bornschier, Volker. 1980a. "Multinational Corporations and Economic Growth: A Cross-national Test of the Decapitalization Thesis." *Journal of Development Economics* 7:191–210.

[9] Controlling for FDI, for initial GDI as a share of GDP, and for initial level of GNP per capita. For significant negative findings in subsamples see Bornschier (1980a).

[10] Controlling also for the logged level of GNP per capita, which has by far the most significant effect (a positive one), especially in the subsample of 88; and controlling for the initial level of Exports as a share of GDP.

[11] For a critical review of the empirical effects of trade and trade dependency variables on growth, see Bornschier and Hartlieb (1981). This research includes also a reanalysis of previous work of other researchers.

————. 1980b. *Multinationale Konzerne, Wirtschaftspolitik und nationale Entwicklung im Weltsystem*. Frankfurt and New York: Campus. [With contributions by other authors.]

————. 1981. "Dependent Industrialization in the World Economy: Some Comments and Results concerning a Recent Debate." *Journal of Conflict Resolution* 25:371–400.

————. 1982. "Kapitalabhängigkeit von Multinationalen und Wirtschaftswachstum im internationalen Vergleich: eine Replik zur Kritik von Erich Weede." *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 34:564–70.

Bornschier, Volker, Christopher Chase-Dunn, and Richard Rubinson. 1980. "Auslandskapital, Wirtschaftswachstum und Ungleichheit: Überblick über die Evidenzen und Reanalyse." Pp. 149–92 in Bornschier 1980b.

Bornschier, Volker, and Otto Hartlieb. 1981. "Weltmarktabhängigkeit und Entwicklung: Überblick über die Evidenzen und Reanalyse." *Bulletin of the Sociological Institute of the University of Zurich,* no. 39, pp. 1–75.

Review Essay: Sociology after MacIntyre

After Virtue: A Study in Moral Theory. By Alasdair MacIntyre. Notre Dame, Ind.: University of Notre Dame Press, 1981. Pp. ix + 252. $15.95 (cloth); $7.95 (paper).

Donald N. Levine
University of Chicago

So inured are we to imputations of crisis in sociology that it must be refreshing to be told that another field is *really* in trouble and indeed that the tools of the sociologist are essential to diagnose and treat it properly. In Alasdair MacIntyre's tractate, the field is moral philosophy; the crisis is the "interminability" and "shrillness" of contemporary moral debate; the diagnostic tools are those of what Benjamin Nelson called a depth-historical sociology of culture; and much of the solution would be to inform moral discourse with a fitting sociological perspective.

Yet our relief is only temporary for, properly understood, the crisis MacIntyre depicts is one whose flip side turns out to involve sociology itself, and the malady in sociology would appear all the more fearsome for not even having been suspected. I find it hard to reach any conclusion other than this: if MacIntyre's argument is valid, most sociological work is misguided or harmful, and a radically new set of research programs should be devised for the field without delay.

It will be tempting for sociologists who read this work to ignore those radical implications, partly because they will be put off by MacIntyre's misleading representation of modern sociology, partly because they will not read the book deeply enough to discover them. Although they will be gratified to find a philosopher who is conversant with some of our major figures and ideas—not to mention one who deals with large questions on such a grand historical scale—they will perhaps take umbrage at three features of his sociological discussion. At least I did.

MacIntyre's sociological foil is Max Weber. He makes Weber's account of bureaucracy central to his own depiction of modern society. Yet he lacks the patience to grasp the complexities of Weber's argument and thus, like so many readers, simplifies Weber's views to the point of distortion. He repeatedly depicts Weber's notion of bureaucracy as based on the principle of cost effectiveness (pp. 24, 82), thereby overlooking the major feature of bureaucracy that Weber highlighted by treating it always

as a form of *authority:* its rationality rests first and foremost on its legitimation through rationally enacted legal norms. True, Weber does write about the instrumental rationality of bureaucratic organization in large-scale societies, but he also makes clear that that presents a type of rationality different from formal rationality, which consists of following the rules. Failure to grasp this leads MacIntyre to the strange interpretation that bureaucratic authority is legitimated only through its instrumental effectiveness, that "bureaucratic authority is nothing other than successful power" (p. 25).

Armed, then, with this peculiar version of Weber's account, MacIntyre sets out to represent modern society in terms of a set of ideal types he calls "characters"—those social roles that embody the central moral definitions of a culture. Foremost among these is the Manager. The Manager is Weber's bureaucrat, but one defined (by MacIntyre) as oriented to nothing but the exercise of manipulative power and guided by the belief that he possesses a stock of universal, lawlike generalizations that constitute the grounds for his claim to expertise. Although the character thus portrayed may well catch some important aspects of modern complex organizations, it has two serious shortcomings. On the one hand, by ignoring the distinction between instrumental and formal rationality, it glosses over the tension between the expert professional and the dutiful bureaucrat, thereby losing sight of what many post-Weberian students of bureaucracy have considered fundamental and of what, indeed, constitutes a major axis of moral debate among the "managers" of modern institutions. Moreover, by reducing all managerial work to manipulative play, the account simply blinds one to the effect of those normative expectations that, in all modern institutions, countervail against raw manipulation.

No less misleading than the way MacIntyre caricatures modern society by portraying his manipulative Manager as its dominant kind of character is his delineation of two other characters who round out his modern social landscape, the Rich Aesthete and the Therapist. What all three have in common is "the obliteration of the distinction between manipulative and non-manipulative social relations" (p. 29). Without denying the significance in modern life of such characters and the amoral orientation they embody, one might plausibly draw a very different portrait of modern society by constructing a set of characters whose roles essentially embody a commitment to moral standards. These could include the Activist and the Judge, not to mention the types that Weber himself considered essential to the modern social order, the Political Leader and the Scholar.

Although MacIntyre's glosses on modern society do not purport to represent a serious engagement with sociological literature on the problem, he does profess such an engagement when mounting his critique of

the character of generalizations in social science. And here, I dare say, he is likely to provoke the loudest howls of protest from our fraternity. In MacIntyre's view, the social sciences aim, above all, to produce universal, lawlike generalizations that form the basis for strong predictive capabilities. Although some social scientists may still espouse such aims, they must, by now, be in a small minority. Surely it is widely known that the record of social scientists as predictors is hard to document and poor, indeed, when it can be documented. Yet it is wrong to assume that this difficulty is not just as evident in the natural sciences. Each of the sources of systematic unpredictability in human affairs identified by MacIntyre has its counterpart in the world of nonhuman phenomena. This point was made with particular emphasis by none other than Max Weber (and his senior contemporary, Georg Simmel). We can predict the evolution of a species, the path of a tornado, or the direction of an avalanche (Weber's own example) no more easily than next year's inflation rate or long-term trends in human population. Surely most social scientists, like natural scientists, understand that, insofar as predictions of naturally occurring events are to be ventured, they are best made on the basis of maximum knowledge of the particulars of a concrete situation rather than on the basis of a few universal generalizations.

MacIntyre's description of the attitudes of social scientists toward their generalizations is no less misleading than his strictures about predictability. He maintains, for instance, that social scientists continue to cling to their generalizations in the face of recognized counterexamples and that their generalizations lack scope modifiers. His evidence for the former point is the persistent credibility of James C. Davies's "J-curve" theory of revolution, despite evidence that would discredit it. Of this it must be said that sociologists of revolution have long since passed beyond the simple correlation of Davies's thesis. Moreover, as the last generation of work by historians of the natural sciences shows, it is quite common in *those* disciplines for widely held generalizations to be maintained long after the appearance of contradictory evidence. On his second point: textbooks in sociological methodology now routinely insist on the provision of scope modifiers for all generalizations. This is not to say that it might not be justifiable, on occasion, to have generalizations of very wide scope indeed. The problem with the law that MacIntyre cites—G. C. Homans's "If the interactions between the members of a group are frequent in the external system, sentiments of liking will grow up between them"—is not that it is so general but that it is patently false (not to mention ungrammatical) and was never widely accepted by sociologists (I hope). But Homans's generalization that by avoiding favoritism a group leader enhances both the solidarity of the group and the leader's own authority has not, to my knowledge, been undermined empirically.

From this part of his analysis, MacIntyre concludes that the social sciences need to "start out afresh" and find their philosophical ancestry, not in the likes of Comte and Mill, but in Machiavelli, with his respect for Fortuna, bitch-goddess of unpredictability. On this point, however, I would find his position better represented by advocating a renewed affiliation with Max Weber who, after all, issued a forthright call for a probabilistic social science. MacIntyre cannot express such advocacy because he finds the contemporary vision of the world predominantly Weberian and sees this vision as one that "depends for its power on its success at disguise and concealment" (p. 103). Which brings us to the heart of his argument, one that in spite of the flaws outlined above, sociologists would do well to consider closely.

I

Surveying the world of contemporary moral debate, MacIntyre finds a plethora of ethical positions and claims, all of which purport to derive from impersonal rational argument but whose oppositions in no way seem amenable to resolution through rational discourse. MacIntyre sees this state of affairs legitimated by what he regards as the prevailing orientation in both contemporary philosophy and public parlance: emotivism. Emotivism is the doctrine that all moral judgments are nothing but direct or indirect expressions of feelings and hence beyond subordination to rational criteria. However, although emotivism rests on a claim that every attempt to provide a rational justification for an objective morality has failed, and thus entails the rejection of any rationally grounded distinction between manipulative and nonmanipulative social relations, its hegemony is challenged in practice by the recurrent pretensions of moralists to a rational defense of their positions. These contradictions are explicable only if they are regarded as dislocated fragments from what was once an integral body of moral thought, in which moral standards were tied quite comfortably to the justification of consensually supported moral beliefs.

For MacIntyre, the great historical agency of that fragmentation was Enlightenment philosophy. The Enlightenment project involved a vain attempt to produce a universal, secular morality based on the sociologically false assumption that human beings possess an individuality that is prior to, and outside of, all social roles. In the wake of the Enlightenment failure, "both the utilitarianism of the middle and late nineteenth century and the analytical philosophy of the middle and late twentieth century are alike unsuccessful attempts to rescue the autonomous moral agent from the predicament in which the failure of the Enlightenment project of providing him with a secular, rational justification for his moral allegiance had left him" (p. 65).

Because modern philosophy is anchored to the fragments of what was once a coherent program, philosophical analysis by itself can in no way penetrate or transcend the problem. What is needed is a social scientific view of the entire culture, including a retrospective view of the integral culture from which these fragments originated. This leads MacIntyre to examine the heroic culture of Homeric society as the matrix of that classical view wherein morality and social structure are linked indissolubly, where the good consists in proper performance of one's social roles through the exercise of relevant virtues. This view found its classic philosophic exposition in Aristotle's formulation of the virtues as the locus of the good, a formulation later cloaked with sacral legitimation through the medieval achievements of Christianity, Judaism, and Islam.

What MacIntyre then proposes as a way out of the contemporary moral Babel is to restore an essentially Aristotelian conception of morality, but one revised in three important respects. It must be stripped of Aristotle's metaphysical biology, one that posits a natural *telos* for man. It must be cured of the ahistorical character of Aristotle's understanding of human nature and further historicized by locating all moral choice within a "narrative" context. Finally it must correct Aristotle's assumption that tragedy consists only of the conflict between moral aspiration and human weakness by incorporating the understanding, portrayed in the tragedies of Sophocles, that the truly tragic conflict is the conflict of good with good. In MacIntyre's final complex articulation, this neo-Aristotelian approach rests on an equation of morality with the possession of virtues, defined by him as the acquired human qualities that enable us to achieve goods that are internal to practice, that sustain us in the quest for the good, and that furnish us with increasing self-knowledge and increasing knowledge of the good. Because of MacIntyre's deep alienation from the contemporary political and social order, he supposes that the kind of characters needed to embody this new moral definition could find a home only in small-scale local communities; thus he closes with the hope that a leader for a modern quest after virtue in these "new dark ages which are already upon us" (p. 245) will take the form of another Saint Benedict.

II

This is not the journal in which to confront the specifics of MacIntyre's account of the history of Western moral philosophy, nor would I presume to do so. But it is pertinent here to ask, What happens if one turns this reconstruction over and looks at what it implies about the history of Western sociology? It would be something like the following.

Surveying the world of contemporary sociological debate, MacIntyre would find a plethora of sociological positions and claims, all of which

purport to be justified on rational grounds but whose oppositions seem in no way resolvable through rational discourse. This state of affairs is buttressed by a doctrine which MacIntyre cites as the sociologists' precise counterpart of emotivism—the doctrine of value neutrality. MacIntyre takes this doctrine to mean not what Weber intended—an injunction to social scientists to refrain from making value judgments in the course of their research—but the belief that sociologists cannot ground any kind of value judgments rationally. By extension, these would have to include value judgments implicit in research programs in sociology, since all such agendas derive ultimately from nonrational assumptions about the worth of those investigations (Weber's "value relevance"). Yet this belief is accompanied by arguments presented in impersonal terms purporting to defend one or another research orientation on rational grounds. These contradictions are explicable only if they are regarded as dislocated fragments from what was once an integral body of social thought, in which the inquiries of "social science" were tied quite comfortably to the justification of consensually supported moral beliefs.

The great historical agency of that fragmentation was Enlightenment social science. The seminal procreator of that project was Thomas Hobbes, who expounded a false philosophy—that human actions are understandable largely as the outcome of atoms in motion, hence subject to the laws of a human mechanics—and thereby introduced the vain attempt to produce universal, lawlike generalizations as the sole legitimate mode of representing them. Following the Enlightenment project, the evolutionisms of Comte, Marx, and Spencer of the middle and late 19th century, and the plethora of behavioral "sciences" in the middle and late 20th century, constitute so many unsuccessful attempts to justify this mode of investigation by appeals to scientific reason.

Sociological analysis, by itself, cannot penetrate or transcend this problem. What is needed is a view of the enterprise that involves a retrospective conceptualization of the philosophical matrix from which these fragments came. The philosophy that Hobbesian science sought to overthrow was basically Aristotelian, one that insisted that the sciences of human action are essentially practical, in the sense of their being oriented to the promotion of human happiness through finding ways to enhance the cultivation of human virtues.

The contemporary solution would thus be to restore an Aristotelian approach to social science, while removing its metaphysical biological basis, its lack of historicity, and its failure to appreciate the pervasiveness of moral conflict. This entails a forthright advocacy of the social sciences as disciplines devoted to understanding the range of human virtues, their sources and modes of development, and the conditions of their flourishing in diverse institutional contexts.

III

In espousing a sociology devoted to the understanding and promotion of human goods and the practices and institutions which embody them, in maintaining that a binding morality cannot be derived from abstract universal calculations but is necessarily related to the institutional realities of a particular historical context, and in suggesting that the fulfillment of social functions is a master clue to the delineation of binding norms, MacIntyre has, in effect, reinvented Durkheim (and, to a considerable extent, George Herbert Mead). What is more, he provides what amount to three telling improvements over a straight Durkheimian approach. (1) Just as MacIntyre faults Aristotle for neglecting what Sophocles represented so vividly—the tragic conflicts that stem from the irreconcilability of diverse human values—so Durkheim's sociology can be corrected by supplementing it with a Weberian stress on the inexorability of conflicts among values. (2) Instead of (I would argue, along with Weber, in addition to) representing human institutions and norms in objective structural terms, MacIntyre argues for the appropriateness of "narrative" reconstructions as a plausible way of representing the vicissitudes of virtues and vices in historical experience. (3) MacIntyre replaces an account of the virtues in terms of behavior in the performance of roles with a stress on the notion of "goods internal to practices" as an essential defining characteristic of virtuous activity and contrasts that with the (typically corrupting) effect of "institutions" with their orientation toward external goods like money, status, and power.

In neglecting Durkheim, MacIntyre missed the opportunity to ally himself with a superbly congenial social theorist. What is more, he missed the possibility of interpreting the whole panorama of modern European history, not as a fall from grace, but as a process of secular differentiation, in which the "fragments" can be seen as meaningful parts of the evolution of a society now organized through sociocultural pluralism. That, of course, would be an attractive possibility only to someone not committed to condemning the entire modern societal order as a dark new age.

In this perspective of what has been called "differentiation theory," MacIntyre's critique loses some of its force—but is still packs plenty of wallop. For if this view can be taken to legitimate a variety of sociological agendas, it does stipulate also that a major direction for social research has been sorely neglected. At the heart of any sociology after MacIntyre must be a research program informed by a neo-Aristotelian, post-Durkheimian vision: one that seeks to identify the social and cultural functions proper to particular historical settings, to delineate the external resources and internal practices needed to realize them, and to show ways of establishing conditions that both sustain us in the quest for the good and

furnish us increasing self-knowledge and increasing knowledge of the good. That this is not an entirely fanciful proposal is implicit, after all, in MacIntyre's allusion to the man whom earlier generations often cited as the "first sociologist," Adam Ferguson. "It is Ferguson's type of sociology," writes MacIntyre, "which is the empirical counterpart of the conceptual account of the virtues which I have given, a sociology which aspires to lay bare the empirical, causal connection between virtues, practices and institutions" (p. 182).

Book Reviews

The Sociology of Virtue: The Political and Social Theories of Georges Sorel. By John L. Stanley. Berkeley and Los Angeles:. University of California Press, 1981. Pp. x + 387. $30.00.

J. S. McClelland
University of Nottingham

Those who know John L. Stanley's previous work on Sorel and who have waited for a fuller treatment of Sorel will not be disappointed. In *The Sociology of Virtue,* Stanley offers a view of most of what Sorel wrote, and he does this for a very good reason. Stanley wants his own seriousness about Sorel to convince us that we should begin to take Sorel seriously. Sorel has been treated too often as a marginal figure in the histories of other people's doctrines. He gets a look-in when revolutionary activism, violence, Marxism, positivism, fascism, or Bergsonianism are being discussed, but he is seldom placed at the center of serious discussion; he always seems to be on the outside looking in. Sorel flirts with other doctrines, or the holders of other doctrines flirt with him. The result has been that Sorel has not been taken seriously *as Sorel*. This diffusion of Sorel has led to the mistaken assumption that, because Sorel seems to be at the center of nothing, there must be no center to Sorel. Stanley wants to demonstrate that there is a center to Sorel, that the center is persistent, and that a view of the entirety of his work is impossible without understanding that center. The center is Sorel's moralism, and the center holds through the early work on the Greeks, the critique of positivism and of Marxism, and is still there when we get to the Sorel that everybody knows, the Sorel of syndicalism and the *Reflections on Violence*.

This view of Sorel means that Stanley has to take *Le Procès de Socrate* (1889) very seriously. He shows convincingly that Socrates was on trial as the representative of those who want to argue us into and out of moral positions without giving us a reason why we should ever feel that we should live our morality. We meet the same Socrates in the socialists, democrats, utopians, and positivists of around 1900 whose view of the world is so anemically evolutionist that they believe that progress will arrive automatically at some time in the future. Nobody needs to lift a finger, because the future has been mapped out by the past. This is especially true of the orthodox Marxists who have to see society simplifying itself into two great opposing camps, the bourgeoisie and the proletariat, so that their own simplistic social science will be capable of explaining society. A simplistic social science has as its analogue the

automaton-proletarian who is fair game for the factory owner and the revolutionary party. The worker has been tamed by the factory system only to become the tame labor force for the politics of a revolution dreamed up by intellectuals who cannot convince the proletariat that they should take part in the revolution. Orthodox Marxists arrive at this impasse because they have misunderstood what a social science is and what it deals with. Marxists try to imitate ideal Laplacian observers who discover how "out there" works. They forget that the real scientists are engineers who make an imprint on reality and who discover the principles according to which reality works while they are in the process of changing reality itself. (In laboratory science, it is the apparatus makers who set the limits to both what "things" can be analyzed and what can be found out about them.) The engineer does not "apply principles." Just as the poet does, he makes something, because he is capable of an act of will. Hence Marxism as "social poetry." Intellectuals, whether theoretical physicists or party theorists, come off badly when compared with the engineers, those who know what it feels like to create. The latter change the world, and their exhilaration contrasts with the sloth of ordinary men. They do not embody goodness conceived of as abstract right or correctness of theory. They practice virtue in a sense that Homer's heroes would have understood.

This summary of part of Stanley's argument does not do justice to the whole book, whose virtue is that it keeps the general argument going while offering rich insights into Sorel's separate concerns. The central theme of Sorel the moralist controls what Stanley has to say about individual themes without ever dominating or schematizing those themes. Methodologically, the book is a model for all such enterprises. The highest compliment one can pay to Stanley is to say that he has not said the last word about Sorel. The evident care he has taken with his bibliography shows that he means us to go away and read more Sorel. Perhaps another book will be written which will modify some of Stanley's judgments, but it will be Stanley's Sorel that will be judged, because for the first time in English we have a Sorel who is now worth the trouble of a serious trial.

The Politics of Social Theory: Habermas, Freud, and the Critique of Positivism. By Russell Keat. Chicago: University of Chicago Press, 1981. Pp. xi + 245. $25.00 (cloth); $7.95 (paper).

Richard Harvey Brown
University of Maryland

Russell Keat is one of the new breed of British social thinkers who is philosophically analytical as well as politically radical. *The Politics of*

Social Theory is an attempt—generally successful, always interesting—
to fuse these two commitments. Keat uses Habermas's theories of the
relations between knowledge and power as a field of discourse on which
to mark his own position. He then assesses Freudian theory and practice
as an example of Habermas's "critical reason" at work.

Keat's intellectual strategy is to distinguish various uses, meanings, or
aspects of key terms in Habermas's texts and then to show that some of
these operate logically in the way Habermas wishes, whereas others do
not. Thus Keat is able to narrow the scope of Habermas's warrantable
claims, specify their proper application, and then, from the residue of
what Keat has cut out or Habermas omitted, construct a defense for a
reformulated positivism that could be integrated with a radical political
commitment.

This volume advances our understanding of the issues and concepts at
stake in the debate on the nature and uses of the human sciences today.
I cannot reproduce here the copious range and prudent rigor of Keat's
rich argumentation, much less state the points and grounds of my dis-
agreements with it. Yet, by defining and distinguishing meanings as they
relate differently to the same premises or claims, by self-reflectively stating
reasons, intentions, limits, and criticisms of his own position, and by
writing with intellectual modesty and great clarity of style, Keat has
produced a work that will enlighten and provoke his critics and admirers
alike.

The telos for Keat is to fuse a Weberian notion of knowledge with a
Marxian image of freedom. He begins by challenging Habermas's central
theses that interests inhere in any given form of knowledge
and, thus, that positivists' claims to value freedom are scientistic mis-
representations. In building his case against Habermas's theory of knowl-
edge, Keat judiciously summarizes its main tenets and then marshalls a
number of arguments against it. These are, among others, that Habermas
uses an inadequate theory of science as the butt of his attacks; that
scientific and moral discourse are two different language games, with
different appropriate criteria for scientific and normative judgments, re-
spectively; that the interest-constitutive theory of knowledge is itself pos-
itivistic (i.e., it has been most used and abused by positivists themselves
in their effort to create their version of a "rational society"); and that a
scientific politics, whether it be based on critical or on orthodox notions
of science, is impossible and probably undesirable.

Keat then turns to Habermas's use of psychoanalysis as an example of
a liberating practice that also expresses an emancipatory theory. Here
again Keat marshalls many ingenious arguments against Habermas's for-
mulation. He argues, for example, that Freud's theories are less critical
and much more deterministic than Habermas's presentation of them. He
constates that freedom from compulsiveness could mean freedom to do
evil guiltlessly, that is, that theories of personal emancipation still have
to treat the problems of evil and of social order. And finally, Keat argues

that even if Freudian theory and practice were both valid and effective, this would not mean that there is any necessary relation between the two. (To demonstrate such a relation logically, Keat argues, one would first need a theory of Freudian *practice*—rather than just a Freudian theory of the psyche—as well as a theory of the relation between theory and practice generally, both of which are sadly lacking.)

In the final section of the volume, Keat reminds us of the complexity of norms in practice and the difficulty of knowing or doing good in lived situations. He concludes that Habermas's conception of freedom and rationality is excessively cognitive and subjectivist and that, therefore, even as a normative ideal it is not a very useful guide for practical discourse or action.

This treatise no doubt will be much criticized, and, indeed, Keat has begun this in his own text. Much more important than the validity of specific points or basic assumptions, however, is the thoroughgoing precision and cogency of his presentation. Keat has enunciated a powerful challenge to critical theorists to refine and strengthen their own positions, and to Weberian radicals to show specifically how their intellectual practices serve emancipatory ends.

The Irony of Liberal Reason. By Thomas A. Spragens, Jr. Chicago: University of Chicago Press, 1981. Pp. xii + 443. $23.00.

Don Herzog
Institute for Advanced Study

Enlightenment liberalism has become something of a straw man in recent decades. Classical liberals, as we have been solemnly and endlessly instructed, were overly optimistic, ahistorical, individualistic, thinly veiled ideologists for the nascent bourgeoisie. And so, the tale goes, we must escape the strangling confines of liberalism to attain, depending on the critic's point of view, either community or Athens. It is a pleasure, then, to find an intelligent study which attempts to take liberalism seriously while criticizing it and seeks "to resuscitate the humane essentials of the liberal tradition, properly understood" (p. viii).

Liberals, Thomas A. Spragens, Jr., contends in *The Irony of Liberal Reason,* worked out a trenchant new epistemology with which they gleefully attacked their Scholastic predecessors. Along with their empiricism came a vibrant new vision of political theory; armed with their shining new method, liberals would dispel the clouds of illusion haunting the unenlightened and generate a moral science capable of certifying liberal norms and saving us from the absurdities of feudal and mercantile abuses. Yet the putative synthesis unraveled, because moral argument could not meet the stern demands of empiricism. And so liberalism, worn away by

internal criticism, committed "conceptual suicide" (p. 183), leaving us two sets of decidedly illiberal heirs. Technocrats like B. F. Skinner eschew talk of values and try to substitute a scientific engineering of malleable man. And "political irrationalists" like Weber and Sartre reduce all norms to mere personal preference, leaving us helpless to explain what is wrong with totalitarianism.

Yet, the argument continues, technocrats and irrationalists alike have retained a fundamental assumption of empiricism, which Spragens names "epistemological manicheanism": anything worthy of the name knowledge must be "absolutely luminous and indubitable" (p. 45); all else is so much garbage, to be ignored. And that strand of empiricism, culminating in the Vienna Circle's positivism, has been subjected to a penetrating critique. The constructive project of the book, then, is to show how new views in epistemology and philosophy of science allow us to escape the dilemmas posed by the illiberal heirs of liberalism.

There is, as Spragens recognizes, a resident danger in painting intellectual history with such broad strokes of the brush. All the quirks and oddities, the puzzles and contradictions, will tend to be smoothed out in the interest of establishing the main theme. That theme, in turn, is stated so abstractly that it is often hard to know what to make of it. It is, at the very least, jarring to find that "the genius of the Panopticon, then, was its concrete embodiment of the Cartesian subject-object relationship and the conversion of this epistemological situation into a political institution," or that "Weber's social scientist is differentiated in an important way from Kant's theoretical ego, however, because he no longer inhabits a Newtonian cosmos" (pp. 120, 252). I never find this sort of talk illuminating, though I daresay many readers will.

Inevitably left behind, as the Enlightenment goes chugging on to its ignominious end, are some misleading and controversial judgments. Spragens quotes the notorious first sentence of Freud's "Project for a Scientific Psychology" (p. 158) without telling the reader that Freud soon abandoned such silliness. And though Spragens has little sympathy for Marxism, Bentham is once again "leather-tongued" (p. 108), and is also as demonic a figure as was ever dreamed by Michel Foucault. But perhaps Hume receives the least sympathetic treatment. His *History of England* is "very favorable toward the Stuarts," with no evidence given to buttress this claim (p. 273). While Spragens seizes triumphantly on Hume's infamous pronouncements on the inability of reason to generate moral distinctions, he has the context wrong: Hume's *Treatise of Human Nature* 3.1.1 is surely an attack on the likes of Samuel Clarke, not on Aristotle or "classical moral rationalism" (p. 225). And I think he miscasts Hume's alleged belief in invariant human nature, ignoring what Hume does say about authentic and important cultural variation and arguing that Hume serves up warmed-over teleology (p. 235).

But we can leave such quibbles aside. There is another danger, to which I think Spragens is somewhat less sensitive. The links forged here

between philosophy and politics are awfully tight. It is as though the foundations of the political world were epistemological, and then bricks were simply added on from there to build a structure determined almost uniquely by the foundation. To employ a better image, I would suggest instead that we think of philosophy and politics as two parts of an enormous web of interconnected beliefs supporting each other, with no foundation at all. That idea still leaves room for links between epistemology and politics, but the links will be a little more tenuous (see here J. W. N. Watkins, *Hobbes's System of Ideas* [London: Hutchinson University Library, 1965], or David Miller, *Philosophy and Ideology in Hume's Political Thought* [Oxford: Clarendon, 1981]).

Far more successful than Spragens's reading of the Enlightenment is his elegant sketch of recent pioneering work in epistemology (see chap. 8, "The Quiet Demise of Liberal Reason"). Lucid and compelling, this discussion succeeds in summarizing a complicated literature without caricaturing it. In the last chapter, Spragens, with undue modesty, attempts a brief sketch of the political implications of the new view. Politics, he argues, can be a "rational enterprise" without using reason to pronounce with certainty on objective moral truths—a view worth applause and attention. The view he sketches is much like Mill's conception of democracy as government by discussion.

It is not quite Mill, though, because Spragens has a warm spot in his heart for Aristotle (e.g., note pp. 193–94). It is odd to find a self-professed critic of the analytic/synthetic dichotomy claiming that politics must be dedicated to the common good "because that is the *telos* of politics by definition" (p. 383). Here some good old-fashioned liberal strictures are called for. Spragens thinks "it is hard to improve here on Aristotle, who observes that the purpose of politics 'can only be the good for man' " (p. 383). In undifferentiated Athens, Aristotle's view was obvious: if politics pervades all spheres of life, where else can man pursue the good? Today it is more problematic. For all its crude empiricism, the liberalism of Hume and Adam Smith provides lovely tools for arguing against the rebirth of Aristotle that Spragens seems to flirt with at the close of an otherwise fine book.

Political Pilgrims: Travels of Western Intellectuals to the Soviet Union, China, and Cuba. By Paul Hollander. Oxford: Oxford University Press, 1981. Pp. xvi + 524. $25.00.

Jeffrey C. Goldfarb
New School for Social Research

The reports, findings, revelations, and memoirs of political pilgrims, Paul Hollander's term for intellectual tourists in a new idealized society, have played and continue to play an important part in modern political culture.

In the 19th century, de Tocqueville's *Democracy in America,* a classic of social theory and research, was a report of one such pilgrim. In the 20th century, the results of pilgrimage, as Hollander has shown extensively, have been decidedly less rewarding. With bewildering repetition, repression has been praised as freedom, injustice as justice, despotism as democracy, inequality as equality, and even genocide as liberation.

Hollander's task is to describe and account for the credulity of Western intellectuals visiting the Soviet Union, China, and Cuba, as well as Vietnam, Albania, and Mozambique. His research topic is richly promising. The research results do not match the promise. *Political Pilgrims* is well-documented. The testimony of a broad array of political pilgrims is reported and analyzed. But the methodological approach to the testimony is seriously wanting.

From Jean-Paul Sartre to Susan Sontag, from Pablo Neruda to Carlos Fuentes, from the Webbs to the Berrigans, from Jerome Davis to Henry Wallace to George McGovern, Western intellectual thinkers, artists, theologians, and politicians have been making remarkably foolish declarations about contemporary socialist societies. Hollander takes these declarations as his data. He documents how slave labor camps have been perceived repeatedly as the most humane form of penal correction, how purge trials have been ignored, how the despot Stalin has been perceived as a "wise and caring" leader, how the destruction of liberty has been identified with the creation of community, and poverty with moral purpose. He points to the relationship between intellectuals and political authorities in the West as the root cause of moral blindness and obtuse political judgment, and to strategies of hospitality in socialist societies as capitalizing on this blindness and obtuseness.

The central drawback of the book is that Hollander does not try to understand the political pilgrims and their judgments but to explain them away through a rather coarse variety of sociological and political reductionism. He too quickly passes judgment, too automatically makes sociological deductions.

Because intellectuals affect estrangement from their immediate political environments, according to Hollander, they find no value in those environments. They turn, therefore, to existing socialism for assurance that their utopian yearnings are realizable. Fools that they are, they find ideals realized in the midst of purgatory. They assume that the enemies of their enemies must be their friends, and in doing so, they prove with their gullibility that the notion of the critical capacities and dispositions of intellectuals is little more than a self-justifying myth. Intellectuals who do not follow the pattern—that is, those who are strongly critical of the socialist orders in ways Hollander approves of, among them sociologists such as Peter Berger, Edward Shils, and Lewis Feuer—are cited by him as scholarly authorities. Others, such as Peter Worsley, Todd Gitlin, and E. J. Hobsbawm, he classifies as intellectuals and dismisses. That both the "authorities" and "intellectuals" face serious problems in their com-

parative observations and judgments is not considered seriously by Hollander.

Hollander sees Julian Huxley's recommendation that "it is no good viewing everything Russian through your own imported atmosphere, for that merely acts as a distorted lens to the facts . . ." as the predisposition which enabled sympathetic visitors "to discard the ballast of their critical faculties with a clear conscience" (p. 104). Yet, such a predisposition could be appreciated equally as the necessary attitude of the objective observer. Indeed, it was de Tocqueville's great accomplishment that he was able to view the America of the early 19th century through American eyes and, only then, add his own judgments.

This is not to say that Huxley adopted the attitude of an objective observer. He, along with other political pilgrims, failed through his acceptance of the official "socialist order." The predisposition Hollander mentions partially explains this. But of greater importance is the fact that the pilgrims too readily viewed socialist realities through their own "imported outlook," not perceiving the distinction between the socialist order and the official account of it. This holds true for both Hollander's "authorities" and his "intellectuals."

Hollander, along with his authorities, views the socialism of the Soviet Union, China, and Cuba as totalitarian. For them, the ruling party, its ideology and leadership, dominates society totally, as is claimed in official propaganda. All social developments are seen as a consequence of totalitarian intention—not only grand foreign and domestic policies but relatively trivial matters such as the reception of foreign intellectual guests, most of whom are unknown. The general disadvantage of this view is that it ignores countervailing social forces, such as those that crystallized in the Poland of 1980. The disadvantage of Hollander's analysis is that the techniques of hospitality are viewed simply as a part of a highly effective ideological campaign, a tool of totalitarian manipulation, while less sinister aspects of the techniques of hospitality are overlooked.

The socialist authorities greet political tourists in much the same way as nonpolitical tourists are greeted around the world, presenting an easily appreciated package (Potemkin villages and all) with political rather than old-world, Mediterranean, or tropical "charm." These tourists are approached, not as sources of commercial income, but as official "delegations." They are accorded the respect and special treatment owed to all official foreign and nonforeign delegations, and the result of this treatment is a distorted view of socialist reality. To maintain that the distortion is always the rational intention is to simplify social life.

Such simplification is fundamental to Hollander's presentation. Hollander considers Norman Mailer's use of literary hyperbole as if it were an analytic statement (pp. 210, 414). He claims deficient identity formation as a major cause of the duplicity of the intellectuals, without documenting the process of that formation. Intellectuals, he maintains, turned to Vietnam for "cathartic relief due to the great (if unjustified)

initial burden of guilt these visitors carried" (p. 272), but he does not analyze the guilt or the process of catharsis. When he analyzes in detail the mass media celebration of China and is dismayed by its inaccuracies (compared with general academic knowledge in the West), he seems to believe that this is characteristic only of the political pilgrimages of mass media and not of mass media celebrations in general (see p. 346). While he rejects the real political logic of the pilgrims when they purposively overlook or deemphasize the problems in the socialist order (see esp. p. 110), throughout the work he uses real political arguments for support of the capitalist order, ridiculing those who are overly concerned with its injustices (see pp. 321, 437).

If Hollander had stuck to the task of documenting and analyzing the political pilgrimages of our day, this could have been an important book. Instead, he engages in rather empty polemics with his subjects, and the result is oversimplification. He, like others who call themselves, or have been called, neoconservatives and new philosophers, is disenchanted with the world of intellectuals and denounces it. In the process, he and they fail to take the intellectuals' perceptions and worlds seriously. There is irony in this. Hollander, like the intellectuals he denounces so vociferously, allows his disenchantment with their world to color his observations. As Hollander shows, they lost their critical faculties when they embarked on their pilgrimage from the "West" to the "East." He, a Hungarian refugee of 1956, who moved from East to West, seems to have lost analytic rigor when faced with their stupidity.

The Rites of Rulers: Ritual in Industrial Society: The Soviet Case. By Christel Lane. Cambridge: Cambridge University Press, 1981. Pp. x + 308. $47.50 (cloth); $14.95 (paper).

Richard Hellie
University of Chicago

Using concepts developed by Clifford Geertz, Victor Turner, S. B. Ortner, and others, Christel Lane, in this important book, *The Rites of Rulers,* poses the question whether the Soviets can consciously create and effectively manipulate symbols and rituals known to be successful in gemeinschaft-type groups and societies to aid in the management of conflict and the formation of consensus in a gesellschaft-type society. Her evidence tends to show that, indeed, ritual can be used as a tool of cultural management to inculcate desired political norms and values and that ritualization of social relations can be manipulated successfully as a means of maintaining social control.

The founders of the Soviet state, who stressed rationality while they mocked the rituals of tsarism, would have been horrified at what has come to pass. Yet, even during their time, today's growth of ritual was planted by the masses in spontaneous demonstrations in celebration of

May Day and the October Revolution. Although Stalin's "cult of person-ality" diminished the early stress on rationality, recent developments, largely the work of "ritual specialists," have been a product of the con-scious creativity of Nikita Khrushchev.

In 1964, Soviet leaders, deciding that ritual could be an ideological weapon, resolved to develop and support what had been started by local initiative. As always, the decision-making process was veiled, but I doubt that decisions were made in an attempt to compensate for the declining growth rate, which had not yet begun. Instead, they were probably related to Khrushchev's 20-year plan which envisaged that communism would be instituted by 1981. In the core of the book, Lane portrays in detail the various forms of ritual that were created and are currently in use: family life cycle, initiation, labor, calendrical cycle, military-patriotic, and mass political rites. Some have been quite successful (such as the initiation rites), others (such as the life and calendrical cycle rites) less so. Most rituals, directed primarily at youth and manual laborers, attempt to incorporate the "three holy Soviet traditions": the 1917 Revolution, World War II, and labor. An important segment of the book contrasts Nazi ritual with Soviet ritual and concludes that, while the former em-phasized mythos, the latter emphasizes logos.

Part of the background of the new and reinforced rituals is Marxism-Leninism, a "political religion," in Lane's terminology. She finds a problem in the fact that, in Soviet symbolism, history often replaces Turner's natural-physiological pole (which arouses basic emotions). This stems partly from the fact that she ignores dialectical and historical materialism and "scientific socialism" (to which the Soviets constantly refer and ma-nipulate to provide both the political legitimacy and the stimulus for dynamic change and innovation that she claims they lack) with their obviously mystical nature and supra-empirical power. Moreover, it is likely that some of the Soviet ritual symbols have considerably more historical depth than she surmises. Although Lane observes correctly that many Soviet rituals are highly syncretic, this may not be as handicapping as she supposes in a multiethnic civilization that historically, beginning with Russian Orthodoxy itself, has been a combination of numerous ele-ments.

It is true that many of the Soviet symbols (hammer and sickle, red kerchief, red banner, red star, eternal flame, blazing torch) have been overused, cheapened, and debased, but I can hardly agree that it is difficult to distinguish signs and symbols, that the symbols have not acquired multivocality. Lane's point is well-taken that the ritual symbols of earlier societies were more concrete and immediate than the Soviet frequent use of words, and that this flow of verbiage may be ironic in light of the fact that ritual was introduced to eliminate a need to talk about things. Nevertheless, this change (plus additional use of sounds, graphics, and light) by the Soviets may well represent an evolutionary advance in the sophisticated use of symbolism and ritual in a "multimedia approach" toward social control. Lane emphasizes that some failings of

Soviet ritual are a result of excessive use of a "model for social relations" rather than of the more traditional "model of social relations," but we must recall that the didactic element is far stronger (and of greater antiquity and consistency) in Russian civilization than, perhaps, in any other. Although the Soviet goal of using the power of ritual symbols to transform individual consciousness into social consciousness, to convert the obligatory into the desirable, by a fusion of the emotional (the individual) and the ideological (the social) is something new, it represents only a variation on long-standing, deliberate attempts by progovernment as well as dissident elements to transform the Russian popular consciousness. Lane concludes that, in spite of some failings, the Soviet ritual specialists have been successful.

As addenda to Lane's book, I might mention that the Soviets have been given high marks for the ritual elements of the 1980 Olympics. A recent Soviet book on military rituals (by Lieutenant general V. D. Serykh, *Voinskie ritualy* [Moscow: Voenizdat, 1981], p. 159) makes a considerable attempt to root contemporary rites in the early years of Soviet power.

This book was used with great success in the University of Chicago freshman "Self, Culture, and Society" course, but its exorbitant price certainly precludes many such adoptions.

Competitive Comrades: Career Incentives and Student Strategies in China. By Susan L. Shirk. Berkeley and Los Angeles: University of California Press, 1982. Pp. xiv + 231. $24.00.

Thomas B. Gold
University of California, Berkeley

Competitive Comrades is not the book Susan Shirk intended to write originally. What was begun in the late 1960s as a study of the role of China's schools in promoting Maoist ideals of equality and cooperation became an examination of the ways in which the structure of educational selection and job promotion fostered the exact opposite form of social relations—competition, alienation, and cynicism.

Although, on one level, this is a detailed study of behavior in Chinese high schools between 1960 and 1966, Shirk uses schools as a microcosm for Chinese society as a whole in order to develop general principles to explain much of behavior in communist systems. Her contention is that revolutionary regimes that attempt the moral transformation of society distribute opportunities to those who best exemplify the desired moral virtues. She coins the term "virtuocracy" to describe such regimes, contrasting them with others that award life chances based on professional or intellectual ability (meritocracy) or ascriptive characteristics (feodocracy).

The use of virtuocratic criteria for educational and occupational selection and promotion assists the leadership in promoting social transfor-

mation, mobilizing the masses for economic development, and contributing to political consolidation and legitimation. Like charisma, virtuocracy is difficult to routinize once a revolutionary movement comes to power. The standards it sets are vague and subjective, thereby open to manipulation by leaders and cheating and sycophancy by the upwardly mobile who know they will be evaluated solely on the basis of observed behavior. The reliance on judgment by peers as well as by outsiders fosters intense competition, rather than unity, among aspirants. Finally, the system generates social conflict through the encouragement of mutually harmful actions. Instead of promoting cooperation and unity, then, virtuocracy elicits adaptive behavior that has alienated people from each other and from the state.

To explain why virtuocracy has such negative consequences, Shirk first examines the opportunity structure that sets the context for student behavior. By 1960, the chances of getting into top high schools and colleges, or of receiving a good job assignment in the city, were quite limited. To improve their chances, students chose one of two basic strategies based on an assessment of their own comparative advantage and the criteria being stressed at any given time: becoming either a political activist or an academic whiz. After setting out this context and choice of strategies, the book describes varieties of interpersonal relations among the students. The effect is far from what was intended: "acrimonious political competition, avoidance of activists, retreat into the private world of friends and family, and disaffection from the regime" (p. 4).

The study is based primarily on in-depth interviews conducted in Hong Kong in 1971 with 31 student refugees (seven females) and three teacher refugees, most of whom were originally from nearby Canton. There were follow-up interviews in 1978, and the author visited China four times between 1971 and 1980. In addition, she utilized official primary sources, unofficial Red Guard muckraking publications, and fiction. In an appendix, Shirk presents a convincing case for this research methodology and her sensitivity to its possible pitfalls. Having spent a year at a Chinese university with students comparable in age and background with those in her sample, I can affirm the validity of her presentation and the reliability of her data.

The strength of this book lies in its multifaceted depiction of the micro-order in China. Shirk does a first-rate job of presenting the context and the thought processes of ambitious young Chinese weighing their chances for upward mobility in a politically volatile society of scarcity. She intersperses quotations from her respondents to substantiate her points, but the fragmented nature of these quotations prevents the reader from gaining a complete picture of any one individual. While preparing this review, I read a remarkable autobiography, *Son of the Revolution* by Liang Heng and Judith Shapiro (New York: Knopf, 1983), which recounts one person's experiences in this system, illustrating and corroborating the picture Shirk has drawn in a more sustained fashion.

Virtuocracy is a challenging concept that needs to be explored and

refined further. What can it tell us about Chinese society that an analysis based on class conflict cannot? Has moral transformation been behind the distribution of life chances? Was the Cultural Revolution an attempt to purify society or primarily a power struggle parading behind moral slogans? How is the amount of virtuocratic criteria weighed in a particular decision? How much of the pervasive alienation in Chinese society today can be traced to the social system of schools (and other institutions) in the 1960s? My own interviews suggest that students were not alienated from the system at that time. Even the Anti-Rightist Movement and Great Leap Forward did not shake their commitment. But the exposure of hypocrisy and corruption and their own experience of it in the social breakdown of the Cultural Revolution are the real sources of their alienation. They may have felt this existed in the early 1960s, but had there been no Cultural Revolution, I would argue, virtuocracy alone would not have produced the disaffected behavior so rampant today.

The Modernization of China. Edited by Gilbert Rozman. New York: Free Press, 1981. Pp. xv + 551. $22.50.

Martin King Whyte
University of Michigan

Several years ago Immanuel Wallerstein declared at an ASA convention that modernization theory was dead. It is apparent, however, that significant segments of our discipline believe, to paraphrase Mark Twain, that reports of this death have been greatly exaggerated. A group of scholars at Princeton University, in particular, continues to "keep the faith." A few years ago these scholars organized an interdisciplinary team to give us *The Modernization of Japan and Russia* (by Cyril Black et al. [New York: Free Press, 1975]), and now they have produced *The Modernization of China* as a companion study. Their work is premised on the notion that many of the criticisms of modernization theory can be dealt with, and that its framework can give us a more satisfactory understanding of modern social change processes than its competitors, such as world-systems theory. To be specific, modernization theory has been taken to task for being ahistorical and using simplistic ideal-type concepts. The Princeton group therefore applies the framework to specific historical developments in Japan, Russia, and now, China. The modernization framework has been faulted also for reifying the nation-state as the unit of analysis and thus ignoring the changing international environment and the manner in which outside forces constrain and distort developments within any one poor country. The Princeton group handles this charge by treating international market forces, imperialist incursions, geopolitical realignments, and other external influences as part of their historical account. Finally, modernization theorists have often been criticized for

confusing Westernization with modernization and assuming that, in the long run, other countries will come to look "just like us." The Princeton scholars argue that, by considering cases arising in a non-Western context and cases that develop through state socialism rather than capitalism, they should be able to discern the core requirements of modernization and some of the institutional alternatives that can fit these requirements.

In their study of China, Gilbert Rozman and his interdisciplinary team of nine scholars adopt a novel periodization, presenting their treatment of modernization trends in two main periods: prior to 1900 and after 1900. They argue that, because the year 1900 and the defeat of the Boxer Rebellion which occurred in that year marked a major shift toward large-scale official efforts at institutional reform, it is more suitable as a breaking point than the dates usually used, namely, 1840, 1911, or 1949. In each of their two major time periods they consider changes taking place in five institutional realms: the international context, political structure, economic structure, social integration, and knowledge and education. Modernization is conceived as an "umbrella term" referring to a set of interrelated modern change processes: increasing international dependence, relative growth in nonagricultural production, a demographic transition, sustained economic growth, bureaucratization, mass political participation (democratic or not), specialization and proliferation of organizations and skills, a more even distribution of incomes, and expansion of education at all levels, to mention only some of the more important. In their study of China they consider whether, and how fast, the various change processes were taking place in different periods and some of the reasons for their success or failure. They make frequent comparisons with the experiences of other countries, particularly those of Japan and Russia.

How convincing and satisfying is the result? To me, the first half of the book is more original and satisfying than the second. The discussion of the years before 1900 is organized around a clear and perennially intriguing question: Why was China, which appeared to be "precocious" in its possession of many modern features, such as a meritocratic bureaucracy, extensive market and contractual relationships, and political unity, less successful than Japan and Russia in modernizing in the 19th and early 20th centuries? The authors consider but eventually place little weight on the two explanations most favored by previous scholars: imperialist attacks on China's sovereignty after 1840 and the complacent "middle kingdom psychology" of China's 19th-century leaders. Instead they argue that trends (or nontrends) already visible in the 18th century made the situation less favorable than in Japan and Russia. In that century, rapid population growth without significant adjustments in the size of the bureaucracy or the tax base left China's rulers less able to extract resources from society than in earlier periods; and, in the 19th century, the spread of literacy, growing urbanization, and other trends that prepared the way for modernization in Japan and Russia were not visible in China. But the problem was also political. Since the existing institutions seemed to be working tolerably well, China's rulers did not

develop a sense of crisis and a commitment to reform until very late in the day. Readers inclined toward other explanations may not find these arguments wholly persuasive, but they will still find the discussion of social trends prior to 1840 intriguing.

The second half of the volume unfortunately lacks such a clear focus. For the most part the authors catalog the changes that were or were not occurring in their various institutional realms after 1900 and interlard this with mention of historical events and trends in the various time periods. Giving coherence to this discussion is admittedly made difficult by the occurrence of major breaks in Chinese politics—the downfall of the Manchu dynasty in 1911, the warlord period up to (roughly) 1927, the Nanking decade of Nationalist rule up to 1937, the anti-Japanese and civil war periods up to 1949, and, finally, the Communist period. There are many interesting observations and insights about Chinese trends and problems, but still the reader does not gain as clear a sense of how modernization was proceeding or what the obstacles were as in the first half of the book.

The second half of the book also makes some of the other problems with modernization theory all too apparent. As treated here, it is not really a theory in the strict sense but an attempt at an empirical generalization: certain specifiable change processes tend to occur together. Some versions of the genre, such as W. Goode's research on family change, attempt to spell out the logic behind these interrelationships, but, in this book and many other treatments, the logic remains implicit. The nature of the framework does not lend itself easily to hypotheses other than those of the kind, "The more industrialized a country, the more urbanized it will tend to be." But since it is acknowledged that the relationships among the various change processes are loose, and that, in each society, there will be "leads and lags," it immediately becomes difficult to evaluate specific changes. Is the fact that China today has a low level of urbanization in relation to its level of economic development "good for modernization" (by saving on funds for urban consumption and social services and by keeping talent in villages) or "bad for modernization" (by interfering with the concentration of creativity and talent that urbanization fosters)? Should one be concerned about the fact that certain aspects of income distribution and the decline of fertility in China resemble those in more developed countries? Also, since functional substitutes are acknowledged as possible, how does one tell whether an attempted alternative development approach is conducive to modernization? For instance, are the many participative devices implemented in Maoist China to be faulted because they do not provide for such things as democratic elections (p. 316) or ways for interest groups to articulate those interests (p. 389), or do they meet whatever imperatives of participation may be entailed by modernization? While one may accept the general change processes of modernization, the descent from the lofty heights of urbanization, bureaucratization, and so forth to judge the implications for modern-

ization of alternative factory payment systems or primary school curricula puts one in much more uncertain territory immediately. So it seems unclear what guidance is provided by this framework for potential modernizers in the Third World.

The failure to make explicit the logical implications for modernization of the many specific changes in each major institutional realm leaves us dependent on the "seat of the pants" judgments of Rozman's team. (It should be noted, though, that two of the members of that team, Cyril Black and Marion Levy, have, in some of their other writings, tried to spell out some of the linkages.) In this situation, informed readers are likely to find many of the arguments presented in the volume less than fully convincing. For instance, it seems clear that the Princeton group considers the Cultural Revolution decade (1966–76) generally a setback for modernization while the ensuing "pragmatic" period has been much more favorable. But it might be noted that, in terms of various indicators, modernization was making headway in the former period, with economic growth proceeding, much progress toward the universalization of schooling, and the beginning of the rapid decline in fertility. There are also reasons to suspect that the shift in policies since 1976 represents more than the requirements of modernization bringing China back to reality. In the highly centralized bureaucratic system created after 1949, the egalitarian measures promoted by Mao Zedong were profoundly threatening to China's bureaucrats and intellectuals, and, with Mao and his radical allies gone from the scene, these threatened groups might be expected to use their power advantage to terminate egalitarian programs and create new privileges. So, should the many incentives and perquisites enjoyed by Chinese elites since 1976 be seen as required by or, at least, conducive to modernization, or should they be seen as representing the victory of "new class" interests?

Perhaps the source of such problems in using modernization theory, as with its ancestor, evolutionary theory, is not that it is incorrect, but that, even if correct, it does not provide the basis for making clear predictions or hypotheses about specifics. Thus, even if the other criticisms of modernization theory can be remedied adequately, as the Princeton group has endeavored to do, it still does not provide a very adequate framework for explaining historical developments in China. At the same time, it should be said that alternative frameworks, such as world-systems theory, have not been very successful in explaining trends in China. Indeed, whether or how to incorporate state socialist societies within the latter framework remains a basic theoretical problem. This leaves modernization theory alive and, as shown by this work, capable of generating useful ideas and insights into historical change processes. If, in the end, the understanding of modern China provided by this framework is not satisfying, this does not mean that we should turn to other available approaches, but that much remains to be done in the task of constructing a satisfactory theoretical account.

The Development of Capitalism in Colonial Indochina (1870–1940). By Martin J. Murray. Berkeley and Los Angeles: University of California Press, 1981. Pp. xii + 685. $34.50.

Bruce Cumings
University of Washington

This is a landmark work. It is one of the few English language studies of colonialism in Indochina and, perhaps, the only one to combine a rigorous analytical framework with a thorough empirical grounding based on extensive exploitation of French archival materials. The framework is Marxist, more specifically, the author uses "the method of historical materialism" (p. ix) to elucidate the development of capitalism in Indochina, and, furthermore, to take this particular case as an instance of "the essential prerequisites for the development of capitalism proper" (p. ix). *The Development of Capitalism in Colonial Indochina (1870–1940)* is, in other words, an ambitious study that hopes to fill out Marx as it fills in the unfortunate paucity of knowledge about colonial Indochina.

Martin J. Murray does not take Marx for granted as he does Wallersteinian world-systems theory which is, for him, an unproblematic conception of the character of colonial rule: an international division of labor in which colonies provide labor and raw materials, local markets for metropolitan goods, and an arena for a restricted industrialization that serves only metropolitan interests. The author makes no secret of his view of this as an entirely one-way affair: asymmetric, exploitative, and ultimately "barbaric" (p. xi)—and not just in Indochina. This is the nature of colonialism generally, he argues, although without the necessary convincing comparisons with other colonial experiences. The other major Asian experience with colonialism, that of Japan in Korea, Taiwan, and Manchuria, is nowhere mentioned (I will return to this).

Murray ranges widely over a period of nearly a century, treating colonial state policies, the emergence of a working class, agrarian social relations, rubber plantations and mines, and, an often neglected but absolutely central process, banking. He devotes more than 50 pages to an analysis of colonial banking, especially the "central financial octopus" (p. 132), the Banque de l'Indochine, which was the sole bank of issue throughout the colonial period and, by the 1930s, the dominant economic force and, in many ways, the dominant political force in Indochina. There is much here that could be compared with, for example, the role of the Japanese-dominated Bank of Korea. Both provisioned credit throughout the colonial realm, both manipulated and overwhelmed private capitalists, and both benefited enormously from the economic crisis of the 1930s. The author also leavens this account with a fascinating portrait of two bank managers, René Thion de la Chaume and Paul Baudouin (pp. 152–54); the latter parlayed his colonial fortunes into major right-wing political influence back home in France. The section on banking is an important, richly rewarding contribution to the literature.

Murray emphasizes that colonially induced development meant that capitalism did not "emerge organically" but was, so to speak, affixed here and there by exogenous forces. Hybrid forms of production mixed the advanced with the primitive or, rather, mixed capitalist forms of production with non- and precapitalist forms. Therefore, saying that all this was capitalist, feudal, or semifeudal misses the point and reduces a rich and complex array of social forces to largely meaningless, formal abstractions (p. 2). Everywhere a "modern," advanced capitalism interacted with a complex, diverse milieu that was itself in transition. Hybrid forms of production gave way to a hybrid social development, so that the colonial working class became a worker-peasant category; when the French were finally booted out of Indochina, they left not a modern working class but a mélange of people suspended between field and factory. Like other colonial powers who introduced industrialization to the colonies, the French ended up with the worst of both worlds: peasants were radicalized by the harsh and exploitative introduction of factory and plantation production, but little or no stability was achieved because peasants were destroyed as a class and yet were not integrated permanently into factory production. Murray presents a variety of statistics on wage rates (pp. 354–58) and working conditions to show that the French mortgaged human capital instead of investing in it, seeking through draconian means to get maximum output in the short run. Here, too, the author makes a major contribution to the literature on the subject.

The long section on agrarian class structure (pp. 374–472) is more familiar, perhaps because there has been so much more scholarship on this aspect of French colonialism. The author accepts the notion that French rule was far worse in its consequences for peasants than the indigenous traditional system, because labor relations moved from "a protective paternalism" to an "increasingly rigid impersonal and contractual connection" (p. 463) in the countryside. Many scholars of Vietnam say this, but I am skeptical of such judgments because they are not borne out in China and Korea, where capitalism or colonialism brought a more regularized form of exploitation that may or may not have been harsher for the victims in its consequences but that also, in dialectical fashion, opened up unheard-of opportunities for its beneficiaries. This does not mean that capitalism is not exploitative, but, rather, that it moves forward in waves of creation and destruction, and a full analysis must account for both.

Limitations of space make it impossible to give Murray's Marxist analysis its due. Readers will find that this is not just a book about Indochina; it is more properly a book about the origins of peripheral capitalism. Murray has read Marx carefully, and he seeks a rigorous Marxist analytical framework. For my taste, much of the analysis lacks the subtlety of Marx's own work. At its best—and this includes several hundred pages—this does not detract from the achievement of this very fine book. At its worst, however, it means that the reader is treated to too much repetition (I never want to hear "expanded reproduction of capital" again in my

life) and to such a mechanical application of categories that classes go around continually doing things, as if moved by some *deus ex machina*, with the goal always being to squeeze the last ounce out of Indochinese natives. Thus, "monopoly capital . . . remained indifferent to both wage levels and the home market" (p. 108); "the fractions of metropolitan capital" were interested in colonial investments (p. 100); landowners, shopkeepers, traders, and "similar agents of colonial rule figured as a supporting class fraction in the particular configuration of the terrain of class domination" (p. 122); and so on. Such mechanistic analysis sometimes substitutes for the terribly difficult problems of locating, say, small entrepreneurs in a general pattern of colonial exploitation. Also, several chapters (e.g., chap. 3) are unnecessarily lengthened by repetitious theoretical sections that merely recapitulate what the author has said already.

A more serious error, in my view, is to assume that the specific features of French colonialism in Indochina define the general features of colonialism everywhere. In spite of qualifications from time to time, this is what Murray thinks. Yet, had he investigated Japanese colonialism in Northeast Asia or British colonialism in Southeast Asia, he would have found an imperialism that was just as brutal as the French (more so in the case of the Japanese, I think) but that also bequeathed skills and an infrastructure that accounts, in part, for the economic dynamism of semi-peripheral Taiwan, Singapore, and the two Koreas. Many are called but few are chosen, Immanuel Wallerstein says, and the successes of this group of the chosen should not surprise a world-systems analyst. But they probably would surprise readers who accept Murray's relatively unrelieved view of colonial exploitation.

These criticisms are minor; the achievement is major. Martin Murray's book will from now on stand like a mighty rock in front of those seeking to justify (or mystify) the character of French rule in Indochina.

Modernization and the Working Class: The Politics of Legitimacy. By Carlos H. Waisman. Austin: University of Texas Press, 1982. Pp. ix + 244. $27.00.

James Petras
State University of New York at Binghamton

Modernization and the Working Class focuses on the relationship between modernization and the legitimacy of the capitalist social order; in particular, it examines the conditions under which the industrial working class accepts the process and consequences of the incorporation of the working class in capitalist society, evaluating the political-deterministic views of Lipset and Bendix as well as the economic deterministic views of Marx. Carlos H. Waisman describes several typologies dealing with stages of political mobilization, the various outcomes of the crises of incorporation,

and their structural and contextual determinants. He posits four sets of outcomes to the "crises of incorporation": accommodation, co-optation, polarization, and exclusion. Waisman then elaborates several typologies relating forms of working-class political action as a prelude to a discussion of several hypotheses dealing with the interplay between elite strategies and worker responses. The study examines three historical cases as illustrations of the latter: Disraelian Britain, Bismarckian Germany, and Peronist Argentina. The author finds that, in capitalist societies which were early industrializers, with few precapitalist classes and in control of their own (and other countries') economic resources, the ruling class tended to be accommodative; in the second and third wave of industrializers, which tended to have precapitalist classes, conditions were more propitious for exclusion or (short-term) co-optation.

The second half of the study is basically an examination of survey data focusing on the Argentine working class in the 1960s. Essentially the study examines several structural features of the working class (skill, old and new workers, etc.) and concludes that deprivation generates reformism, while large concentrations of workers who are vulnerable to crisis can produce a revolutionary effect.

Waisman has made a significant theoretical contribution by integrating world-systemic variables (timing of industrialization, dependency) and social structure features (class and state) with an analysis of collective political action (elite and working-class). The analysis is complex and dense, but the result is worth the effort.

Two issues require clarification. First, Waisman's discussion of capitalist accommodation and worker legitimation among the first industrializers is based on the assumption of continued world supremacy and expansion. What happens to this consensus when global positions begin to decline seriously and economic depression undermines the available resources? The general notion that the correlation between modernization and working-class acceptance of capitalism is curvilinear may be called seriously into question. The second and related issue is that there is insufficient explanation for the changes from one set of political activities to another (reformist to revolutionary or back)—in Waisman's work, the relations in the formative period become, in a sense, frozen over time. The shifts from exclusion (1870–90), to accommodation (1890–1933), to exclusion (1933–45), to accommodation (1946–) in Germany, for example, require an analysis of the impact of the imperialist wars, the Great Depression, and the postwar economic boom. The impact of these world-historic processes (part of the global environmental factors that Waisman identifies) needs to be theorized or integrated into the theory of capitalist development because they affect the crucial variables of legitimacy and elite/working-class relations.

Waisman's conception of capitalist development overlooks its tendencies to systemic crises and contraction. Instead, his study tends to flatten out these tendencies and to look at "modernization" as a finished and

self-sustaining operation. The deepening crises of contemporary capitalism (early and late industrializers) certainly makes it evident that economic and social regression are very much part of our environment. These objections aside, this is an important and rigorously executed theoretical and historical study that merits careful reading.

The Class Struggle in the Ancient Greek World: From the Archaic Age to the Arab Conquests. By G. E. M. de Ste. Croix. Ithaca, N.Y: Cornell University Press, 1982. Pp. xi + 732. $49.50.

Joseph M. Bryant
University of Toronto

The relevance of Marxism for the study of antiquity should no longer be questioned after this immense work of scholarship by a distinguished classicist. Combining sociological acumen with solid historical craftsmanship, G. E. M. de Ste. Croix provides a new and insightful account of the major institutions, social groups, and historical developments of some 1,400 years of Greco-Roman civilization (from the Archaic age to the Arab conquests), and the result, *The Class Struggle in the Ancient Greek World,* is not only one of the finest studies on classical society to date but also a major contribution to historical sociology.

The hallmark of Ste. Croix's approach is class analysis, which he modifies in two respects. First, he regards the standard definition of "class" (those sharing a common relationship with the means of production) as too narrow since it fails to convey that classes involve "above all things exploitation" (p. 69), that is, "the appropriation of part of the product of the labor of others" (p. 43). He argues that Marx himself stressed the fact of exploitation, as evidenced by his thesis that "the specific form in which unpaid surplus labor is pumped out of the direct producers . . . reveals the innermost secret, the hidden foundation of the entire social structure" (p. 51). Second, because he believes that class struggle (the singular is deliberate) has been a constant feature of all complex societies, he wants to deny that the term necessarily implies collective action, conflict, or class-consciousness. For Ste. Croix, class struggle exists whenever there is exploitation, that is, the appropriation of surplus. While this conflation of terms seems confusing and unnecessary, it explains why Ste. Croix incorporates "class struggle" in the title of a book concerned principally with exploitation rather than struggle.

Ste. Croix's analysis begins with the fundamental problem facing all complex societies: that of extracting a sufficient surplus from the primary producers so as to give others the "leisure" for politics, art, and so on. Given antiquity's primitive technology, the only solution for the dominant classes was to extract high rates of surplus from the lower orders. This

they did directly through the use of unfree labor—chiefly slaves, but also helots (serfs) and debt-bondsmen—and indirectly (primarily in the Roman period) by tightening the screws on the humbler free through taxation, military conscription, and forced labor services.

With the majority of the population deriving its livelihood from the soil (trade and manufacture were typically small-scale), land and unfree labor constituted the principal means of production. The free citizenry basically consisted of a minority of large landowners and a majority of medium-to-small peasant farmers—the ancients themselves stressed this dualism by speaking of the rich and poor, the few and the many. On the bottom were large numbers of the unfree, whose characterization as "animate tools" by Aristotle well describes their standing. Ste. Croix demonstrates that unfree labor was the major source of surplus for the propertied, a fact enabling the elite to create the political and cultural achievements we still admire. With the entire superstructure of antiquity thus presupposing the existence of dependent labor, Greece and Rome were clearly "slave economies," even though the largest share of the total production was usually created by small peasants (an unknown percentage of whom also owned a slave or two).

Ste. Croix next turns to the problem of social change, and again class relations are shown to be decisive. He provides explanatory accounts of the rise of the tyrants, the flowering and destruction of Greek democracy (which was not as widespread a political form as the book conveys), and Rome's transition from republic to empire. He concludes with a fascinating analysis of the Roman Empire's fall, emphasizing both the change in the form of exploitation (from slavery to serfdom as the humbler free became *coloni* bound to the soil) and its intensification due to an expansion in the "nonproductive" segments of the population: the army, the imperial bureaucracy, and the Christian Church.

The book's most evident shortcoming involves Ste. Croix's denigration of status as an explanatory concept (Weber and M. I. Finley are the main targets). The difficulty is that the forms of open factional violence or *stasis* which fill the pages of Herodotus, Thucydides, Aristotle, and others do not quite correspond to Ste. Croix's exploitation model. Open struggles in antiquity were not for the most part between unfree direct producers and their exploiters but among factions within the aristocracy and between rich and poor citizens (with citizenship an ascribed status). Thus, there existed two distinct forms of struggle: exploitation of the unfree by the propertied, and *stasis* among citizens, which Aristotle attributed to the pursuit of *kerdos* (material gain) and *timē* (honor). In fact, after dismissing several scholars who note the distinction between the master-slave and the rich-poor relationship, Ste. Croix himself yields to the evidence by speaking of "class struggle on the political plane" (i.e., within the citizenry), a process separate from the exploitation of the unfree. Moreover, with "status" ignored, he occasionally strains the meaning of "class," as when he regards the poetry of Theognis (6th century B.C.?), which depicts

the hereditary aristocracy giving way to an order increasingly based on wealth rather than blood, as revealing "bitter class [*sic*] struggle" (p. 279). Such a characterization hardly fits when the contending groups were separated by the quality of their blood rather than by their economic standing.

In sum, Ste. Croix has established the indispensability of class analysis but not its self-sufficiency; hence the need for a historical sociology combining class and status. I can find no better authority for this than Marx himself, who fully appreciated the importance of antiquity's manifold gradations of social rank, and who, when he noted that citizens avoided certain economic activities as "corruption" fit only for metics, freedmen, and so forth, actually anticipated Weber's insight that status often restricts economic forces. Marx also stated that Rome's class struggles (and the same applies to Greece) "took place only within a privileged minority, between the free rich and the free poor, while the great productive mass of the population, the slaves, formed the purely passive pedestal for these conflicts" (p. 61). Ste. Croix's eagerness to legitimize "class" compels him to characterize such statements by Marx as "trivial," an unnecessary action since the consideration of status would not invalidate the substance of his research.

This short review cannot do justice to the encyclopedic range and empirical richness of Ste. Croix's scholarship: on topics ranging from the subjugation of women to the attitudes of Jesus and the early church fathers toward property and slavery, from military affairs to ideological disputes, this book will long serve as an invaluable guide to the ancient world. An affordable paperback edition is therefore imperative, for this work belongs in the libraries of all students of the human condition.

The Social Production of Art. By Janet Wolff. New York: St. Martin's Press, 1981. Pp. x + 196. $22.50.

Wendy Griswold
University of Chicago

> Caught in that sensual music all neglect
> Monuments of unageing intellect. . . .
> And therefore I have sailed the seas and come
> To the holy city of Byzantium.
>
> [W. B. YEATS, "Sailing to Byzantium"]

Yeats portrayed the aspiration of an old poet, dismayed by the flux and fecundity of the world, to sever his connections with material reality ("fastened to a dying animal") and travel to the realm of artifice, becoming a mechanical, golden bird whose song is unpolluted by history or desire.

The prevalent conception of the artist, from the Romantic period to the present, is that of a person who has successfully completed Yeats's longed-for journey. His art is a matter of sending postcards from Byzantium to the folks back home, who slog through the muck of generation.

The sociology of art attempts to correct this notion of the artist as removed from social and historical experience. Two complementary traditions draw our attention. One, the more theoretically animated, locates artistic creativity in an ideological network and points out how the artist, far from transcending prevalent categories of thought, reproduces them. The other, more empirically oriented, tradition explores the institutional and organizational contexts of artistic production. The theoretical approach is favored by Europeans, the institutional approach by Americans. It is to Janet Wolff's great credit that she attempts to combine the two approaches in *The Social Production of Art*, taking stock of what has been accomplished in the sociology of art and assessing the points at which more work is needed. Wolff, a lecturer in sociology at the University of Leeds, is herself more firmly planted in the theoretical tradition, having previously written *Hermeneutic Philosophy and the Sociology of Art* (London: Routledge & Kegan Paul, 1975). And one of the many uses her new book will have is that of a broad introduction to the latest British and Continental work in the sociology of art and Marxist aesthetics. But Wolff is conversant with developments on both sides of the Atlantic, and she brings all parties to the table to negotiate their common and conflicting visions.

(The institutional approach has also produced a major stocktaking volume recently, Howard S. Becker's *Art Worlds* [Berkeley and Los Angeles: University of California Press, 1982]. Both books are far more than subdisciplinary navel gazing; rather, they are timely consolidations of contemporary thinking, from which new theoretical and research developments will surely emerge. I suspect they signal a certain maturity of the sociology of art, a passage out of the time when the field regarded itself as a mongrel yapping at the heels of the Muses.)

Wolff asserts that art is "the complex construction of a number of real, historical factors" (p. 1), and she organizes her book around a series of interdependent but analytically distinct starting points for exploring the impact of these factors on artistic production. These starting points include, first, the mutual involvement of subject (artist) and social, political, and economic structures. She makes the case, not new but well-stated here, that the concept of the alienated artist, aloof from social connections, is itself a historically specific image founded in the breakdown of patronage and the artists' subsequent dependence on an impersonal market, another manifestation of capitalism's tendency to separate producer from consumer. In this respect she maintains that artistic creation is no different from other forms of work; it just looks different insofar as its creative aspects are somewhat less obscured. Following Anthony Giddens's conception of the duality of structure—structures being both constraining

and enabling—Wolff argues that artistic creation is both determined and an arena for situated choices, which constitute freedom. But this freedom is a far cry from the Romantic image of artistic autonomy.

Other starting points for the consideration of art as social production include the collective construction of the preconditions, execution, and reception of art; the relationship between art and ideology (which she defines agnostically as the structured and systematic relationship between ideas and material conditions); the problem of the impact of art on social relations—affirmative as Marcuse feared or alienating and provocative as Brecht hoped; the role of the audience recreating meaning through interpretation; and the place of the author, who, she claims, is decentered but not dead. Devoting a chapter to each of these issues, Wolff offers a perceptive and critical reading of the chief theorists, along with suggestions regarding their ultimate contribution to a total theory of the social production of art. For example, she nicely picks her way through two contrasting hermeneutic positions, that of E. D. Hirsch advocating the valid interpretation of authorial intentions, and that of Hans-Georg Gadamer claiming the inevitability of reinterpretation, "a fusion of horizons," that the prejudiced reader imposes on the author's text. While agreeing with the latter that meaning is always "produced" on the spot, insofar as "reading is always from the point of view of the historical and existential present of the reader," she nevertheless maintains that authorial meaning has priority status insofar as it, though effectively dead, "has historically informed the present reading of the text." And she goes on to draw the methodological import of her synthesis: "A sociology of literature in general, of course, would incorporate original meaning (and its construction), mediation of that meaning through, for example, a series of critics, and meaning attached to the work by any new reader, as well as the interrelations between these" (p. 103).

One might take issue with Wolff's enthusiasm for blaming artistic marginality, actual or perceived, on advanced capitalism. It is simply not correct to state that "before this period [19th century], artists and writers were well integrated into the social structures in which they worked" (p. 11). Some, indeed, were ensconced in guilds or systems of patronage, but others were regarded as marginal and, as Plato emphasized, dangerous. Under a statute of Elizabeth I, actors were grouped with vagabonds and other social outcasts.

More important, one might wish that she had worked her sensitive and sensible insights into a unified theory of artistic production, one amenable to proposition formulation and the generation of researchable problems. One might wish she had followed up on her tantalizing hints about the relationship between an unborn theory of artistic production and an unborn Marxist-feminist theory. But these are requests for a book Janet Wolff has not yet written. The book she has written is an indispensable gloss to the most complex current thinking in the sociology of art. She has shown how the structure of Byzantium must be understood as interacting with the golden bird's song in the process of mutual creation.

Art Worlds. By Howard S. Becker. Berkeley and Los Angeles: University of California Press, 1982. Pp. xiv + 392. $25.00.

Michael S. Kimmel
Rutgers University

What makes art art? Since the Renaissance, art has been understood as a work of exceptional beauty, produced by an exceptionally gifted individual. We know artists by their genius, art by its timeless beauty. Academic art history abounds with effusive prose heralding the work of the genius; one recent biography describes an artist's "unwavering commitment to his personal vision, which enriches our consciousness of our fellow men and heightens our awe for the ever-increasing power of the mighty impulses that enabled him to give us a close view of life's vital forces" (Seymour Slive, *Frans Hals,* 2 vols. [London: Phaidon, 1970]). This view of art is itself an ideology, a social construct, as much a product of its time as is the work it describes. And, like other ideologies which masquerade as timeless truths, it leads to mystification of the work.

Howard Becker's superb new book, *Art Worlds,* elaborates a systematic sociology of art that pierces the veneer of traditional perspectives on art and thereby demystifies the works themselves. It is, of course, a fitting topic for Becker, who has maintained a career as an art practitioner—as both a jazz musician and a professional photographer—that has paralleled his respected career in sociology. *Art Worlds* represents the fusion of these two careers, presenting what might be called a "labeling theory of art."

Art Worlds challenges the tenets of traditional aesthetics—the artist as genius and the beautiful work as expression and embodiment of genius. In their place, Becker begins with a simple epigram: art is the work that some people do. It requires an "extensive division of labor" (p. 13) and "elaborate cooperation" (p. 28) among many people: artists, patrons, critics, gallery owners, librarians, jazz club owners, teachers, audiences, and even those who make the tools of the trade (ink, canvas, paints, instruments). These different individuals together constitute an art world, "the network of people whose cooperative activity, organized via their joint knowledge of conventional means of doing things, produces the kind of art works the art world is noted for" (p. x). Their works are "not the products of individual makers, 'artists,' who possess a rare and special gift. They are, rather, joint products of all the people who cooperate via an art world's characteristic conventions to bring works like that into existence" (p. 35). "Talent" is a social construct; art is "what an art world ratifies as art" (p. 156). This is the essence of Becker's labeling theory of art. So much for genius and beauty.

Becker's main task is to explore the constituent elements of the art world. An art world is more than just a network of interacting people; it is also a set of conventions (artistic work is as normative as it is creative), a number of material and personnel resources mobilized for artistic production, and a distribution network. Works of art, he argues, conform

733

to both the materials at hand and the norms of the distribution network; social and material forces structure the limits of creative genius. "Artists produce what the distribution system can and will carry" (p. 129). Moreover, art worlds also contain normative standards of taste (administered by tastemakers) and are themselves subject to larger social and political contexts.

The state, for example, influences works of art by codifying property rights (art works are commodities in a market) as well as intervening through patronage or censorship. "All artists depend on the state," Becker writes, "and their work embodies that dependence" (p. 191). Similarly, editing is cast as a question of survival—what of the original work will last? "We assess the reputations of whole genres, styles, periods, and countries on the basis of choices made by all sorts of people about whom we know little or nothing" (p. 225).

When Becker confronts the thorny problem of changes in art styles, though, his analysis becomes somewhat disquieting. Change in art, he argues, is a social process, requiring shifts in the constituent elements of the art world. And yet he also claims that these shifts, "the structure of artistic revolutions," are initiated by individuals, acting as individuals: "Change takes place . . . because artists whose work does not fit and who thus stand outside the existing systems attempt to start new ones and because established artists exploit their attractiveness to the existing system to force it to handle work they do which does not fit" (p. 136). Here it appears that change depends on the assertion of individual will against the full sociological weight of the art world. What is disturbing is that such voluntaristic origins of change ultimately lead to a theory of nonchange, of conservative stability, of equilibrium. Art and artists often support the status quo because art worlds themselves define it.

The art world, then, is the universe of artistic production. It sets standards, determines aesthetic values, organizes production, distribution, and consumption. It situates the artist and the work in society; it "creates a history" (p. 346). It is not transcendent beauty, a vital unity with the infinite, or an expression of universal, timeless emotions that make an art work last. Quite the contrary, Becker claims. Simply put, "artistic work lasts when it has an organizational basis that preserves and protects it" (p. 350).

Alas, that puts it a bit too simply. In a sense, this sociology of art is not really about art at all; Becker himself notes that his is more "a sociology of occupations applied to artistic work" (p. xi). But surely something remains of a great work of art after it is placed in its art world, after all the social conditions of its production are explicated. Surely there is something in *War and Peace,* the *Jupiter* Symphony, or *Guernica* that speaks to us perhaps not because of the art world but despite it.

Sociology of art requires a sensitive balance between the social conditions of the production of art works and the power of the works themselves. For something remains of those epochal works that pushed beyond the paradigmatic limits of their art worlds and explored the uncharted

universe, something perhaps asociological. Becker takes us a long way toward redressing the balance away from the effusions of genius and beauty; they must, however, be incorporated into the analysis rather than excluded. Nonetheless, *Art Worlds* should become a standard book in the sociology of culture. It is the clearest, most articulate, elegant, and coherent sociology of art to appear in a long time.

The Sociology of Opera. By Rosanne Martorella. South Hadley, Mass.: J. F. Bergin, 1982. Pp. xi + 228. $24.95.

Stephen R. Couch
Pennsylvania State University, Schuylkill

During the latter half of the 19th century, composer Richard Wagner set out to do the impossible: create a unified art form through a thorough integration of music, poetry, and drama. He failed but in the process provided us with a series of magnificent operas, greatly enriching the musical world.

In *The Sociology of Opera,* Rosanne Martorella also attempted an impossible task: writing a moderately sized work which would describe and analyze the social organization of opera production, the economic and social environment within which opera is produced, and the effect of those factors on operatic style. To accomplish this task would take volumes, as well as a theoretical integration of the work of scholars from a widely diverse cross section of sociological subdisciplines. However, while this book does not accomplish this monumental job, it provides a very worthwhile introduction to the world of opera and a basically sound and interesting analysis of the impact of social and (especially) economic forces on the social organization and production of opera in the United States.

The book begins with a discussion of the social nature of opera and the aims of the research: ". . . this study shows the effects of socioeconomic conditions on opera, specifically with regard to styles of performance, the organization of production, and the content of productions. It shows that the roles of performers have been shaped by these foregoing exigencies" (pp. 7–8). This introduction is followed by a chapter on the history of opera in Europe, locating the art form and the emergence of the roles of virtuoso singer, the middle- and upper-class audience, the composer-performer, and the conductor within the context of a capitalist market system. "Increasingly, these roles have become professionalized; there has been a growing interdependence between artists' specialties, their status, aesthetic norms, and the structure of the market" (p. 37).

Then the setting shifts to the contemporary United States. In a chapter entitled "The Economics of Opera," the financial workings of opera companies are presented clearly, as is the constant economic crisis faced by opera houses (especially the largest). The following chapter links the

economics of opera with repertoire decisions, arguing convincingly that market forces (and, more generally, patronage) are determining factors in decisions about which operas to program. In the opera market, every performance brings a financial loss. Unlike profit-making industries, the more produced, the greater the loss. This makes it essential that, in order to minimize losses, the operas produced satisfy a middle-class audience as well as wealthy patrons. The result: very conservative programming, in which opera war-horses are trundled out year after year, dressed in new wrappings and surrounded with new scenery, but underneath, the same old mares.

The remainder of the book discusses the impact of this rationalization process on the roles of the major participants in the opera network: the management and the star performers. For its part, the role of management has grown in importance along with the growth of opera as big business, the need to rationalize to avoid losses, and the entrance of government into direct arts patronage. Managers' unenviable task of working in a money-losing organization, having to balance artistic aims with enormous financial difficulties, and having to satisfy the demands of a multitude of unions, temperamental singers, and flamboyant conductors, necessitates that successful managers rule with a "Prussian hand in an Italian glove" (p. 124). Conflict, while institutionalized, is ever present. "The divergent nature of occupational identities and perspectives manifests itself in the constant disruption of the collective unit" (p. 147).

Indeed, in considering the star performers, it becomes very apparent that a contradiction which is part of all organizations is especially marked in opera companies: that of the individual versus the collective. Virtuoso singers are trained to be solo stars whose goals include individual success and status and whose occupational milieu contains a cult of the individual virtuoso. Yet opera is a collective product par excellence, requiring the subordination of individual egos to the collective endeavor of producing great art. Martorella is sensitive to this contradiction and to its importance and, while outlining mechanisms whereby much of the resulting conflict has been institutionalized, recognizes correctly that this important contradiction cannot be resolved. Yet increasingly, because of the economic realities and resulting rationalization, the goals of the collective are stressed.

The Sociology of Opera offers convincing arguments for its main point: that large-scale socioeconomic changes, mediated through patronage and combined with the requirements of producing the artistic product, have greatly influenced the nature of that product and the social organization of its producers. At the same time, there are some problems. The book is not comprehensive. One wishes for fuller comparative treatment of Western Opera houses with Eastern European ones, where the market is less "free" and private patronage less important (at base, is rationalization a function of the market system or of factors common to a variety of economic forms?). The book also sees opera from the top down, concentrating on general managers, star performers, and conductors and giving less attention to the more numerous but lower-status players in

the chorus, the backstage, the pit, and the office. The emphasis is also on the largest companies. In addition, certain concepts might have been clarified more fully (e.g., does "style" refer to compositional style, production style, or repertoire choice?). Finally, at times, the writing style is somewhat deliberate, more Wagnerian than Mozartian.

Nevertheless, Martorella has produced an interesting book which provides a good sociological analysis of a fascinating artistic institution. She shows a commendable ability to see where the social structure of opera is similar to that of other kinds of organizations and where it is distinct, and then to analyze it accordingly. The book provides a very good base from which further work in the sociological analysis of artistic institutions can proceed.

SoHo: The Artist in the City. By Charles R. Simpson. Chicago: University of Chicago Press, 1981. Pp. xi + 276. $20.00.

Sally Ridgeway
Adelphi University

SoHo: The Artist in the City is an excellent work on both the life-styles of contemporary artists and on the evolution of SoHo as an urban community of artists and art-related institutions. Charles R. Simpson has contributed substantively to the literatures of both sociology of art and that of community studies and urban planning.

Simpson describes the lives of contemporary artists in New York City, including material on their personal relationships, their family lives, their working networks and affiliations, their problems of success and failure, and he discusses, as well, the art market and its members. He traces with a wealth of detail the social history of SoHo as a community, showing how it was discovered, and, in particular, he chronicles various political issues rising from both the establishment of the community and attempts to assure its survival. One such political fight that engaged the artists involved a bypass highway which would have destroyed the area as a community. Simpson suggests that the conception of SoHo as a status community is central to an understanding of artist life-styles and of contemporary urban planning problems.

Simpson develops several ideas around his discussions of artists and the politics of urban planning. In discussing the success and failure of the artists, he makes some major contributions to understanding art as an occupation. Our society supports a myth of artists as romantic, individualistic, and expressive, engaged in nonroutinized, unalienated labor and rebelling against bourgeois values. The chapter on the unsuccessful artist shows that the further these artists were from a successful career, the greater their commitment to this artistic identity and the greater their need to act out its presumed bohemian characteristics. The need to assert

this identity is obviously important, Simpson feels, otherwise the failed artists might have had to acknowledge that they were no different than those they left behind in the suburbs. Simpson makes a poignant point about the unsuccessful male artist's tendency to see the wife "as the bourgeois upon whom he relies and whom he incidentally seeks to scandalize" (p. 196).

Successful artists, however, have little need to assert a bohemian lifestyle. Here Simpson describes another important pattern—the rationalization of the artists' careers. While the unsuccessful artists are acting out a form of the romantic, self-expressive bohemian life, successful artists—like their middle-class, professional, working counterparts—go to work every day, albeit in a studio, but with equipment, with neat and organized tools and surroundings, and with staff. These artists break for lunch and are tired when they finish work at the end of the day. Their creativity lies in conceptualizing the art work; implementation is hard work, often drudgery, with pressure and deadlines. The bar life, the wild parties, and the romance of celebrity status are not visible here. Other writings on working and successful artists make a similar point about the bourgeois quality of their lives and their disparaging attitude toward bohemianism.

As a community study, Simpson's documentation of SoHo's development is invaluable. SoHo, itself, is today a residential area for artists, an important part of the national and international art market with established galleries and other art market businesses, as well as a thriving area for boutiques, restaurants, and entertainment. Simpson comments that, as a community, SoHo contributes to limiting the "destructive consequences of anomic drift and urban pathology" (p. 238). Discovered by artists who wanted cheap and spacious living spaces, SoHo changed from a dying industrial area to prime real estate as the artists carved out, not just their homes from the factories, but a viable, exciting urban community. But Simpson points out how this process of "gentrification" made SoHo attractive to the affluent middle class and thus susceptible to real estate exploitation. The artists provided labor for this real estate market, clearing and settling the territory, making it safe for others. These artists, now competing with nonartists for what has become preeminently desirable but very expensive living space, are gradually being forced out of their community; thus, the area may be denuded of the very qualities which made it an attractive place to live. "The tide of artistic, subcultural and small business vitality could recede from SoHo, leaving it an empty shell" (p. 243).

There are many other ideas developed within this book that bear noting: for example, the theme of the suburbs as affecting perceptions of self, work, family, and life-styles; particular problems facing women artists; the role of the artist's wife; discussion of the artists' involvement in their childrens' education, in particular, their demand that individualism as a central value be fostered in school; the detailing of the artists' political organizations in the community; the history of these organizations them-

selves; the relationship of diversity and residential involvement to a community's survival. Brief, insightful comments are frequent in this book. For example, Simpson notes how easily the artists used their political organization for personal goals, because they were "culturally disposed to recognize the 'patron' aspects of a relationship when they saw them" (p. 166). Such observations add to the richness and density of this text and to its overall significance.

From Bauhaus to Our House. By Tom Wolfe. New York: Farrar, Straus & Giroux. Pp. 143. $10.95.

Albert Bergesen
University of Arizona

When Leland Stanford, Jr., died in 1884, Leland, Sr., wanted to build a memorial and enlisted the services of famed landscape architect Frederick Law Olmsted, who had designed New York's Central Park and the Berkeley campus. Leland, Sr., wanted a campus of his own in Palo Alto. Olmsted wanted something like Berkeley, with buildings interwoven among trees, streams, and rolling hills. He wanted Stanford further up in the foothills. Leland, Sr., wanted his memorial university on flatland to better realize its essence as a monument to his son. The client wanted a memorial. The architect wanted a naturalistic park. The client got his way. Stanford University was built on flatland, complete with a large triumphal arch, Memorial Arch, and a Memorial Court, a Memorial Church, and a monumental symmetrical courtyard, Inner Quad. I do not know for sure, but it would seem that Olmsted got his way on at least one thing: the Mausoleum is off among the trees, to the right, as you drive up Palm Drive.

For Tom Wolfe in *From Bauhaus to Our House* that is the way client-architect relations should be: tell them what you want. You are paying; you decide. He suggests, however, that things have changed. Today there would be an international competition, a panel of judges composed of a lot of university architecture professors, and a winning design. Some sort of glass box or austere modern building would be built. Think of William James Hall at Harvard. This might not be so bad if most clients did not hate the austere box that became known as the International Style. For Wolfe, both the modern style and the relative independence of the architect began with the post–World War I European architects who formed autonomous "art compounds" like the Bauhaus or the Dutch group De Stijl and were committed to producing, as he puts it, "worker-housing cubes" and a no-frills "nonbourgeois architecture." Thus, they used no ornaments, no facades, no columns, no pitched roofs, no color, no luxurious materials like granite, marble, limestone, or red brick, but only

the bare essentials: concrete, wood, steel, glass, or stucco buildings with flat roofs, sheer facades, thin walls, and colored white, beige, and grey, with an occasional spot of black for accent. "The inner structure, the machine-made parts, the mechanical rectangles, the modern *soul* of the building, completely free of applied decoration" (p. 26). Form follows function. Modern architecture. The glass box. Walk downtown. It is omnipresent.

As if that were not bad enough, says Wolfe, the leaders of this movement emigrated to the United States and dominated American architecture. What began as worker housing in Germany and Holland became the predominant form of corporate American architecture. Worker housing as the New York skyline. Ironic. They not only arrived, argues Wolfe, but became ensconced at the leading universities: Gropius was made head of the Harvard School of Architecture; Moholy-Nagy started the New Bauhaus, later the Chicago Institute of Design; Josef Albers started a rural Bauhaus at Black Mountain College and, in 1950, moved to Yale as head of fine arts instruction; and Mies van der Rohe became dean of architecture at the Armour Institute of Chicago, which merged with the Lewis Institute to become the Illinois Institute of Technology. From Wolfe's point of view, they arrived and conquered. The International Style, as austere modern architecture came to be called, became the standard, and deviation was treated as just that: deviance. There remains a general disdain for nonmodern architecture, although there seems to be some tiring of the glass-box syndrome. There are murmurings of a postmodern architecture. This, however, seems more rhetorical than real, bound up with the writings of Robert Venturi and a lot of talk about the naturalness of Las Vegas or McDonald's golden arches. In short, there seems to be no alternative movement on the horizon other than a pinch of ornamentation here and there under the somewhat inflated label of postmodernism.

For Wolfe the biggest problem was that these Europeans had intimidated Americans—our "colonial complex," he calls it—and given us an architecture unfitting the boisterous imperial position of America at the height of its capitalist glory. We do not need austere, functional worker housing but something more dramatic, decorative, and triumphal to mirror America's place as the hegemonic—or declining hegemonic—state. He contends further that the workers do not even like this kind of functional housing. The only ones who benefit from all this are the architects in their isolated compounds, expounding this and that theory of the ideal architecture.

Then there is the question, What are we to make of Wolfe himself? John Gregory Dunne recently called Wolfe an anthropologist, and I think he was partly right. Some used to say that Goffman was something of a great social observer, watching and reporting all that interaction. But Goffman was really more the theorist—perhaps the American contributor to substantive theory, as opposed to the writings about others which often pass as theory. Maybe social archaeologist is a better label for Wolfe

since he digs up and uncovers social dynamics we do not want brought to consciousness. Wolfe informs us, through sarcasm and with values that often run counter to our liberalism. But, to focus on his conservatism and reactionary nationalism and miss the ironies and contradictions he unveils is a serious mistake.

Wolfe always raises interesting questions. We may not like his answers, or even his questions, but there are things that need to be explained. Why is modern art getting more abstract, or why did the glass-box paradigm take hold in architecture, or how could a socially conscious architectural community concerned with humanitarian values generate such a cold, austere vision of modern housing? You can always tell that his questions have struck some kind of nerve by the inevitable intellectual stonewalling by his reviewers. For such a purportedly lightweight critic, who always misunderstands art, architecture, or Manhattan parties, there is certainly a lot of huffing and puffing in those reviews. So what does the sociologist get from Wolfe? Many things, obviously. But at the heart of what Wolfe knows intuitively are the ironies and contradictions of modern life, whether they are seen in the appearance of members of the Black Panther party at a chic Manhattan cocktail party or in a socialist dream of proletarian housing becoming the totem of 20th-century capitalism.

There are aspects of lived life, dealt with more often in literature than in sociology but real nonetheless, that Wolfe, in his heavy-handed way, brings to our attention. We seem so inept at dealing with human realities dramatized in literature that maybe the slapstick humor of Wolfe is about all we can digest. I fear that if he wrote more subtly there would be little chance we could pick up what he was saying; we have enough trouble with the belly laughter in which he encases his observations now, without his packaging them in a more refined format.

Patterns of Technological Innovation. By Devendra Sahal. Reading, Mass.: Addison-Wesley Publishing Co., 1981. Pp. xvi + 381. $34.50.

John R. Kimberly
University of Pennsylvania

Patterns of Technological Innovation is, in a sense, a curious book to be reviewed in *AJS*. Written by an economist, the book sets for itself the immodest task of developing the basis for a science of technology. Furthermore, the author, Devendra Sahal, addresses the book explicitly to scholars in three fields: economics, engineering, and public and private policy. Why, then, *AJS?* Is there anything here for sociologists?

The answer to the latter question is most certainly yes. In a period of increasing concern about this country's resolve and ability to maintain its central position in the global economic pecking order, the questions

about technological progress raised in this book are particularly timely. And perhaps of greater immediate relevance, the perspectives developed by the author should motivate sociologists interested in problems of technological change and innovation to reexamine some of their more cherished and time-honored premises.

My initial reaction to the book was highly negative. I did not like the author's approach to defining innovation, I did not like what I felt were needless and sophomoric digressions on theory, and I did not like the occasional arrogance in tone. Yet, as I became involved in the book, I found myself agreeing frequently enough with the author's points that I now believe any sociologist working in the general area of technological change and innovation could benefit from a careful reading of it.

Let me cite two examples. One is Sahal's emphasis on the importance of processes and context in influencing patterns of technological advance. Much of the research on innovation has focused on particular items, "innovations," and has examined rates of their adoption. Explanations for differing rates of adoption have been sought either in the characteristics of the innovations themselves (e.g., cost, risk, divisibility, etc.) or in the characteristics of the adopting individuals or social systems or in both. This focus has created an overly rigid conception of what the "innovation" is. Sahal argues persuasively that innovations are modified as user experience accumulates. His argument, if taken seriously, has significant methodological implications for the "innovation adoption" research tradition. At the very least, one would want to understand more fully the evolutionary context of the innovation itself and the process of its development before examining adoption patterns. Certainly examining adoption in vacuo would lose its appeal.

Another example is Sahal's view of diffusion. Based on analysis of a number of case studies, he takes aim at the widely held notion that the S curve is an adequate representation of the process. He urges a close examination of the relationship between new and old techniques and argues, at least implicitly, that the process of diffusion is both more complex and more incremental than might be inferred given existing conceptualizations of the process.

Many of the positions taken in the book are, in my view, highly debatable, and I found myself frequently parting conceptual company with the author. His distinction between "major" and "minor" innovations muddies the theoretical waters. He tends to see more continuity and cumulation in the evolution of technology than I believe exist. And sociologists will most certainly question his lack of attention to analysis of factors which influence whether and how technological knowledge gets used. How, for example, do we account for the *way* in which the automobile has developed?

I doubt that many readers will come away from this book believing that the author has succeeded in his objective of developing a theory of technological progress. But scattered through 317 pages of text are enough

good ideas to make an investment in careful reading surprisingly worthwhile.

Computers and Politics: High Technology in American Local Governments. By James N. Danziger, William H. Dutton, Rob Kling, and Kenneth L. Kraemer. New York: Columbia University Press, 1982. Pp. xv + 280. $25.00.

Martin Greenberger
University of California, Los Angeles

The computer is unremittingly logical and a master of control. It would appear to present local governments with the means to rationalize decisions, streamline administration, and strengthen authority—an executive's delight. At one time, use of the computer seemed almost certain to gravitate to the upper levels of the bureaucratic structure. Indeed, that was an assumption made by the authors of *Computers and Politics*—James Danziger, William Dutton, Rob Kling and Kenneth Kraemer—when setting out on their research project nine years ago.

But top officials in local governments have not found the application of computer technology to overall organizational objectives that simple. The computer's success as a management tool has been spotty at best. Part of the reason has to do with the technical nature of the computer. It falls more in the province of the expert and specialist than that of the manager and politician. The authors assumed, back in 1974, that the computer would be shifting power, not only in the direction of higher management but also toward the technical elite—those possessing the special skills and sophistication required for effective use of the technology. Both assumptions had to be reconsidered.

Computers and Politics discusses the political and organizational effects of computer technology in local governments, using results from a study of 42 cities. The authors raise two fundamental questions of central interest to political scientists accustomed to asking, Who governs, and to what purpose? (1) Who controls computer technology? (2) Whose interests does the technology serve? Not government leaders and the public interest, answer the authors on the basis of their investigations. Not the technocrats either, at least not in any total sense. Who, then?

The researchers found that control and application of the computer in local governments is fragmented, paralleling the plurality characteristic of democratic politics generally. Control is rarely concentrated. Different perspectives on how the computer should be used are evident even among computer professionals, depending on whether they are associated with a central computing facility or one of the using agencies. Computing decisions are not made unilaterally. They emerge from the same kind of

743

political maneuvering, debate, and confrontation of desires observed throughout the processes of democratic government.

Does this mean that the various groups within the administrative structure and influence spectrum take turns in benefiting from computing decisions? No, say the authors. Their central finding is that the dominant coalition within a government is the one most likely to benefit all the time, almost irrespective of the issue or decision. It is the interests of those who already have the power that are often best served by computing. Thus, computing reinforces prevailing values and influence, prompting the authors to call it a "conservative technology and not an instrument for change or redistribution" (p. 231). To protect the values and interests of groups excluded from the dominant elite, they recommend a "democratization of computing [that] might improve the responsiveness of local governments to wider legitimate public interests without reducing the capacity of executives to manage the government" (p. 235).

In arguing for greater decentralization of control in the management of computer resources, the investigators are very much in step with the tempo of the technology. Personal computing is a major growth area in the computer industry today. So is distributed computing. These technological advances provide individual users and departments with access to significant computing resources locally—resources formerly available only from central operations—without the bureaucracy that central control implies.

Decentralization of computing could help governments provide better services to constituents, an area the investigators find reason to criticize. "Our analyses have shown that automated applications are not now widely used to improve the quantity and quality of direct services from local government to the public" (p. 242). In general, the book's commentary on the use of computers by local governments—"prevasive problems . . . deep undercurrent of dissatisfaction with many aspects of computing among departmental users . . . shortcomings and even failures that characterize a substantial number of more sophisticated applications" (p. 243)—is not encouraging, but neither is it surprising. What one wonders is how governments have fared in applying computers, compared with industry and other sectors of U.S. society. What factors make it so difficult and problematic for governments, especially, to reorganize, innovate, and otherwise take full advantage of what the technology has to offer? That is the subject for another book, of course, although there are clues in the interesting case material provided by this one.

The authors pull no punches in presenting the case material. For instance: "Thus Brockton is an example of a sterile milieu for the growth of computing—a partisan political-administrative environment with little extracommunity or community support for computing, where professionalism is not a dominant presence, where factionalism precludes strong leadership, and where those concerned have little understanding of the complexity and needs involved in nurturing an effective computer package" (p. 68).

The subject of *Computers and Politics* is not an easy one to treat or review. One must try to separate out how much of what was observed (e.g., in Eugene, Oregon, and Orange County, California [pp. 240–41]) was a peculiar function of computer technology and how much was connected more specifically to the use of analysis, statistics, or other supposedly objective grounds for policymaking. How much is the ability to use technology effectively itself part of the very source of power that the technology is found to reinforce? Can one discover in the exceptions—in the successful uses of computers by departments not part of the dominant coalition (pp. 189–90)—reasons for attributing *nonconservative* possibilities to the computer as a potential instrument for change? Such questions linger.

In documenting and analyzing this difficult subject, the authors have provided readers with a thoughtful, well-developed thesis to consider, a thesis consciously at variance with early ideas on the role of computing in government. Future researchers will want to study this thesis and the evidence presented with care.

Medicine and the Reign of Technology. By Stanley Reiser. New York: Cambridge University Press, 1978. Pp. xi + 317. $14.95.

Allan Mazur
Syracuse University

To determine the nature of an illness, 18th-century doctors relied on the patient's description of symptoms, on his appearance, and, more rarely, on manual examination of the patient's body. In *Medicine and the Reign of Technology,* Stanley Reiser gives an excellent account of changes in this diagnostic model produced by developments in medical technology over the past two centuries. The patient has become a distant object of study, examined through the media of instrumentation, technicians, and laboratory tests.

The stethoscope was one of the first steps on the road to direct data collection from the patient's body, allowing the physician to bypass often unreliable verbal accounts of symptoms. It was invented in 1816 by a French doctor named Laennec during his examination of a young woman for heart disorder. He wanted to place his ear to her chest to listen to the heart, but her youth and sex deterred him. "Rolling some sheets of paper into a cylinder, he placed one end on the patient's chest and the other next to his ear. Clear and distinct heart sounds emerged" (p. 25). With an improved instrument, Laennec compared the chest sounds of healthy and ill people and was able to relate the differences to anatomical faults he found during autopsies.

Success in hearing the inside of the body encouraged the development of instruments for looking at it as well. With the ophthalmoscope one

could see inside the living eye; other scopes probed the body cavities. Microscopes showed the fine details of tissues as well as the tiny creatures that live inside us. The supreme visual device was the X-ray, showing bony structure through skin made transparent.

Instruments showed body function as well as structure, measuring lung capacity, heart rate, temperature, and blood pressure. With the addition of chemical testing of body fluids, physicians could diagnose and chart the course of an illness almost without talking to the patient.

According to Reiser, all of this technology promoted specialization, not only among doctors but also among laboratories that ran tests and technicians who manned instruments. These specialists processed the patient as if he were a specimen rather than a person, the object of scientific tests, measurements, and probes.

Perhaps Reiser is too quick to blame the new technology for the loss of a humane relationship between doctor and patient. Similar shifts from personal to impersonal relations have occurred in many activities besides medicine, including some where there has been little overt change in the level of technology. Consider the changed relations between merchandiser and customer if one compares the old corner store with the modern supermarket, or the increasingly impersonal relations between professor and student as class sizes have grown larger. In these cases, the number of clients processed in a given time period has increased, forcing interactions to be brief, routine, and formal. This heightened velocity of clients moving through the system reflects changes in organization more than changes in technology, although these changes can reinforce one another.

It is unfortunate that Reiser's history is limited to diagnostic technology, for there are occasional hints of an important interaction between developments in diagnostic technology and therapeutic technology. After all, what good does it do to differentiate disease A from disease B unless you treat them differently? The appeal of a diagnostic technique increases if there is an accompanying therapy. Reiser mentions briefly that the thermometer was adopted for clinical practice at about the time that fever-reducing drugs became widely used (p. 120), and that microscopic counts of red blood cells, indicating anemia, were used in connection with iron therapy which increased the red cell count (p. 132). Although he discusses the X-ray only as a diagnostic tool, much early enthusiasm was also focused on its therapeutic power, with important social effects. For example, in the first decades of this century, the X-ray was used to diagnose "enlarged thymus" among children, a condition then thought to contribute to sudden infant death but now considered innocuous. When this "condition" was found—it was very common—the X-ray became the treatment of choice and was used to shrink the thymus tissue. Thousands of children were subjected to this combination of diagnosis and treatment, and though it had no effect on the rate of sudden infant death, many developed thyroid cancer as adults as the result of the X-ray treatment. Since Reiser has made such a nice start in tracing the social effects of

medical technology, I hope he continues on some of these untouched issues.

The Ceremonial Order of the Clinic: Parents, Doctors and Medical Bureaucracies. By P. M. Strong. London: Routledge & Kegan Paul, 1979. Pp. xiii + 267. $25.00.

Dorothy H. Broom
Australian National University

Many medical sociologists are curious about what really goes on between doctors and patients. *The Ceremonial Order of the Clinic* gives us some answers. It reports on observations made over three years of 1,120 pediatric consultations. Most of these occurred in a city in Scotland, but about 100 were done in the United States. The consultations were mainly between doctors and child patients and their mothers, but occasionally other medical personnel (e.g., therapists) conducted or observed the meetings, and, in a few cases, someone other than the mother accompanied the child. Observations were made in at least 12 distinct medical settings. Yet overarching this "considerable variety" in systems, settings, participants, and problems, one distinctive "format" of encounters which characterized the vast majority of the consultations is identified by P. M. Strong. It is this ceremonial order, namely, the bureaucratic format, to which most of the book is devoted. The author says he has set out "to describe its nature, its origins, the methods by which it was sustained and the actual fit between this outward show and the matters which it clothed" (p. x). For the most part, he relies on qualitative treatment of the data, and the book contains many accounts of how particular consultations were conducted. The approach to social research clearly owes much to Goffman, and Strong acknowledges his intellectual debt in the opening pages.

Such work would not be expected to yield "findings" in the conventional sense. Instead, the reader gradually acquires a sense of some of the tacit rules that guide these encounters, of how role relationships were defined and maintained, of mechanisms preserving the positions of authority, and of strategies for the management of deviance. In the middle five chapters, an image emerges from the descriptions and excerpts of conversations. In the manner of one of the modern diagnostic devices that produces a series of cross sections of the body from which a three-dimensional image can be constructed, Strong goes over and over the typical format, but in a slightly different location or from a slightly different perspective each time, gradually building up a composite picture of what kinds of encounters these are and how they are conducted. To organize and explain the data, Strong relies on the kinds of penetrating distinctions (e.g., "natural versus social frameworks") and useful insights (e.g., the "rule of

irrelevance") that we have come to expect from the Goffman approach. It does not always pay off; for example, he introduces a "hot/cold medicine" distinction (p. 107) without explaining it, thus confusing and distracting me, and then never uses the distinction again. But on the whole, the consultations come to life, and researchers who are interested in face-to-face interaction and in medical encounters will find that Strong has supplied much valuable material.

Despite its inherently interesting data, the book is unnecessarily difficult. Because I had enjoyed several of Strong's articles and found them useful (esp. A. G. Davis and P. M. Strong, "Aren't Children Wonderful?" in *The Sociology of the National Health Service,* ed. Margaret Stacey, Sociological Review Monograph no. 22 [Staffordshire: University of Keele, 1976]), I was surprised to have so much trouble with this monograph. Translating Goffman's insights into anything remotely approaching a theoretical framework (let alone a theory) is notoriously difficult, and Strong does not quite claim to be doing that, so it is hardly fair to criticize him for failing. But the considerable space spent relating this research to Goffman's work bears little fruit. I would rather Strong had devoted more time and care to devising and presenting his own strategy and structure for the book; neither of these is ever spelled out clearly. For example, I would have appreciated a brief, side-by-side presentation of each of the formats, perhaps in a table that contrasted their specific qualities. As it is, the three formats other than the bureaucratic (charity, clinical, and private) each get about four pages placed variously throughout the book, but reviewing these short discussions plus the equally brief treatment of "the concept of role format" in chapter 1 did not enable me to construct the table for myself. Since formats are the focus of most of the book, that seems an important shortcoming, although I found an older version of what I wanted in an article (P. M. Strong and A. G. Davis, "Roles, Role Formats and Medical Encounters," *Sociological Review* 25 [1977]: 775–800). Only one chapter (chap. 8) supplies any preview of what to expect in the pages that follow, but even that is an enumeration of what is to come, not why these elements are included or how they are related. The omission of any discussion of the intellectual strategy of the book (what is to happen or how) makes it a chore to read. It is like getting a jigsaw puzzle with no picture on the box lid.

My specific complaints include a tendency toward repetition and cumbersome expression. For me, the two short critiques of Parsons are no advance on numerous previous efforts. Similarly, the brief attack on what he calls "role theorists" (who are not named) did not tell me anything new about the concept of role or help me understand this book, and none of these discussions was central to Strong's material. Strong has done better by his own argument elsewhere (Strong and Davis, "Roles, Role Formats and Medical Encounters"). The subtitle claims that the book is about "medical bureaucracies," but in fact there is almost nothing about the organizations themselves. There is no mention of similar work on gynecology (Joan Emerson, "Behavior in Private Places," in *Patterns of Com-*

municative Behavior, ed. H. P. Dreitzel [London: Macmillan, 1970]) or of the important research on pediatric consultations (B. M. Korsch and V. F. Negrete, "Doctor-Patient Communication," *Scientific American* 272 [August 1972]: 66–74). The oversight is serious since this research also used direct observation, and Korsch and Negrete discuss the issue of how doctors and patients perceive their interaction, which Strong does rarely.

I was sorry not to see any explicit analysis of gender, either of doctors or of parents. Strong apparently observed over 150 consultations with several different women doctors, yet he says nothing about whether their behavior differed from that of the men, even though this question has puzzled medical sociologists for some time and feminist health advocates have assumed that a major difficulty for female patients is male doctors. Furthermore, he points out differences between the ways fathers and mothers behave in consultations, but he makes nothing of the fact that fathers are also *men.* This weakens his explanation, since, for example, he suggests that fathers are able to be more critical (of clinics, medicine, or doctors) than mothers because they are "outsiders," but he fails to observe that men also, typically, have more power than their wives.

Overall, this is an uneven book. The general discussions at the beginning and end left me unsatisfied, but the substance in the middle—the accounts of consultations that build up a picture of a typical mode of interaction—is lively, interesting, and worthwhile. For advanced undergraduates, graduate students, and researchers in medical sociology, the book could be a useful resource, although Strong has made it a lot more difficult than it should be. He describes one of his own chapters as having "a rather jumbled and somewhat tentative character" (p. 183), which is, I am afraid, my impression of the book as a whole. Our medical colleagues probably will not be willing to struggle with it. I think it would be twice as good if it were half as long.

The Limits of Professional Power: National Health Care in the Federal Republic of Germany. By Deborah A. Stone. Chicago: University of Chicago Press, 1980. Pp. xii + 212. $18.50.

Odin W. Anderson
University of Chicago and *University of Wisconsin—Madison*

The thesis and intended emphasis of *The Limits of Professional Power* are that the source of power of the medical profession is rooted in the political character of a society and is an artifact of that society. The author, Deborah A. Stone, is critical of the prevailing notion that the source of power of the profession is inherent in its technical expertise and the bargaining power flowing from this exclusive attribute. In a very good political analysis of the relationships among the medical profession, the

749

sickness insurance societies to which the government delegates the administration of national health insurance, and the pervasive regulatory influence of the government itself in the Federal Republic of Germany, she tries to support her view that the power of the medical profession is not inherent in its expertise but is solely a function of what a particular political system will allow it to be and do.

I do not believe the author succeeds in supporting her thesis, but, no matter, she makes a distinct contribution to knowledge about the politics of national health insurance and the nature of its relationship with the medical profession and with the government which, over time, has established the legal framework in which the medical profession functions. Even so, the medical profession is no supine entity which simply reacts to the political framework. It is very proactive everywhere in liberal-democratic countries, and even in Communist countries the profession is powerful as the source of expert knowledge for formulation of health service programs.

It seems that the author did not distinguish between two major aspects believed by the medical profession to constitute professionalism. The more the medical profession operates in an open market, the more it believes that control over methods, sources, and amounts of payment and the discretionary prerogatives to diagnose and treat are equally inherent in professional characteristics. As countries adopt national health insurance, or even voluntary health insurance, the profession then begins to differentiate between pure control over methods, sources, and amounts of payment and the profoundly professional prerogative of maintaining control over diagnoses and therapy. The latter then becomes the bastion of the profession, and mundane working conditions and arrangements more and more become functions that society and government believe should not be controlled by the medical profession. Even so, the medical profession continues to bargain quite effectively over methods and amounts of payment, and the delegated power to diagnose and treat, within the resources made available, remains intact, at least so far.

The author is an American and a political scientist at the Massachusetts Institute of Technology. She contrasts (and mentions one or two similarities) the American extremely open health services delivery situation and physicians' greater professional freedom with the situation of the German nonhospital physicians, and she is impressed by the relativeness of the environment in which German physicians practice. Since the turn of the century, and starting with Bismarck's famed introduction of compulsory health insurance for workers in 1883, those covered by health insurance have been organized by the government into sickness insurance societies that administer payments to physicians and hospitals, and the physicians have been organized, also by the government, into corporate bodies so that the two collectivities act as countervailing forces.

Over time, more and more rules and regulations have been applied to both entities: to benefits and premiums of the sickness insurance societies and to methods and amounts of payment and monitoring of physician

practice profiles to separate and penalize the high prescribers in the medical corporations.

Up to the mid-seventies the expenditures of the West German national health insurance (paid for largely by payroll deductions) were going up at a slow but steady rate, but then they suddenly spurted up to a 17% annual increase. Crisis! A Commission on Cost Containment made up of representative interests was established by the government to monitor the system and establish cost constraints of one kind or another, from professional standards review by committees of physicians, analogous to the American Professional Standards Review Organization, to budget caps. Still the government was reluctant to impose expenditure ceilings and continued to mandate increases in premiums.

In spite of all these actions the author concludes that, in general, there is no strong evidence that economic monitoring has had much impact on the volume of physicians' services (outside of the hospital and paid for by fee-for-service). As a group, physicians have been able to "neutralize the system" by shifting the standards of professional norms upward.

The author, then, has produced an excellent case study of the politics of health insurance and the medical profession which has relevance to this country because of the incipient greater structuring of the medical profession and insurance agencies, but she has not proved her thesis that the power of the medical profession is a function of the political system. In the Federal Republic of Germany, as elsewhere among the liberal democracies (and perhaps even in Communist countries with developed economies), the inherent power of the profession transcends the political characteristics of each country. The profession's power to diagnose and treat, and its attendant cost implications, remains intact. Perhaps her thesis will prove true, in the long run, if countries really retrench their health services expenditures. But, before then, maybe the public attitude toward how much doctoring it wants will have changed since it would seem that the physicians' power is largely a function of what the public believes the physicians can do to assuage suffering and postpone dying. These perceptions have a powerful influence on how political systems control physicians, and they transcend differences in political systems.

The Holocaust and the German Elite: Genocide and National Suicide in Germany, 1871–1945. By Rainer C. Baum. Totowa, N.J.: Rowman & Littlefield, 1981. Pp. ix + 374. $25.00.

Frederick D. Weil
University of Chicago

Weber expected a sociological account of a historical event to do two main things: to explain it causally and to render its meaning comprehensible. Whether a sociological interpretation of the significance of the Holocaust,

the mass murder of European Jews by the Nazi regime, is possible now or in the future is problematic, but a causal explanation of "how it could have happened" is possible and needs further development. The literature on Nazism and the Holocaust is so vast that even specialists often hesitate to say that a new work contains new ideas, but it seems to me that Rainer Baum's book, *The Holocaust and the German Elite,* is a solid, if uneven, contribution toward a sociological explanation.

An adequate explanation presupposes that the question has been posed correctly, and Baum shows good sense in beginning with Hannah Arendt's notion of the "banality of evil," not because the crime itself was banal but because the motivation for most of those involved was bureaucratic, *sine ira et studio* (without hatred or passion). Most of the best accounts of mass "authoritarian personality" or of the breakdown of German ideology stop short of attributing the Holocaust primarily to these factors, although the internal logic of such formulations suggests this conclusion, and some authors seem to have been seduced (recently, Fred Weinstein, *The Dynamics of Nazism: Leadership, Ideology, and the Holocaust* [New York: Academic Press, 1980]). Rather, historical research has tended to disconfirm the notion that the masses of German people, the German elites, or the executioners themselves were converted to some kind of bloody tribalism. Bureaucratic crime itself is the phenomenon which must be explained, writes Baum (p. viii), and he rejects not only the passions of hate, greed, and so forth as adequate causes but also the effects of charisma, totalitarian compulsion, or the destruction of the *Gleichschaltung* (intermediate organizations). Even passive or disorganized resistance, if it had constituted an appeal to shared values, would have hindered the widespread and highly organized positive participation in bureaucratic crime, as it in fact did when the murders of retarded and handicapped people became known. Baum also denies, correctly I think, that the murder of Jews or even anti-Semitism was central to the goals or appeal of National Socialism (although he irritatingly persists in calling it a "sideshow" in relation to the "main show" of economic recovery and drive for empire). Nazi success before 1933 was proportionate to economic crisis and the constellation of other parties, and the available evidence about public opinion in the Third Reich does not suggest much support for policies of extreme persecution (see Marlis G. Steinert, *Hitler's War and the Germans* [Athens: Ohio University Press, 1977])—indeed, the regime tried to keep the death squads and death camps secret; and of course, the Holocaust only hindered economic recovery and the war effort. Baum's position is less clear on the theory that the murder of Jews was necessary to tie participants to the regime by implicating them in crime, or to prepare the German people for a larger murder of the Slavs in the drive for lebensraum.

Baum attempts, instead, to account for the Holocaust by explaining why there was no effective resistance among the elites who, he says, were the only ones in a position to do something about it. To this end, he develops a theory of "moral indifference" and devotes the bulk of his book to explaining the phenomenon and showing how it arose in German

social and political history. I am convinced that, as far as it goes, this is the correct line of explanation (see my "The Imperfectly Mastered Past: Anti-Semitism in West Germany since the Holocaust," *New German Critique* 20 [Spring/Summer 1980]: 135–53; coincidentally, I also used the expression "moral indifference" to explain the same thing [p. 137]). However, there are certain problems in Baum's account, even at this stage. First, he argues so vigorously against all other positive motivations that he ends by citing Sebastian Haffner's personality study of Hitler, asserting that "genocide served the private need of Hitler, a need he indulged in only after he realized that the big show, the drive for empire, was doomed" (p. 65): this would seem to bring us back to the authoritarian personality, after all, as the motive force behind the Holocaust. Second, why the Jews? There is no discussion of the nature of German anti-Semitism (although he criticizes another author for treating it as unproblematic [p. 44]), of Germans' images of Jews, or of German-Jewish relations: why were some minority groups, rather than others, singled out by the Nazis for destruction? Third, he excludes the general population and nonelite organizations from his discussion of moral indifference with only the sketchiest of reasons (pp. 10–11), some of them not even empirically correct (he says hardly anyone knew any Jews, but in a 1949 survey, 75% of West Germans said they had known Jews). His implication is that the nonelites were in no position to do anything about the Holocaust, but his own evidence shows that they could have (pp. viii, 111–12), as we have seen above: moral indifference among the general population is problematic and was important.

Baum's account of the genesis of the political culture of the elites, which included moral indifference, is also largely sound and shows an admirable command of the literature. His model is basically a three-step process: an underlying value dissensus in German society encouraged the creation of "one-dimensional" elites (e.g., wealthy industrialists with no political power and no prestige), whose struggles for power helped generate uncontrollable conflict, especially when the masses became mobilized. This is a variation of the theory of status anxiety, and it contrasts especially with Lord Balfour's description of the English, whose "political machinery presupposes a people so fundamentally at one that they can safely afford to bicker" (quoted in Giovanni Sartori, *Parties and Party Systems* [Cambridge: Cambridge University Press, 1976], p. 16). The model is set in the context of the history of German nation building, industrialization, and constitutional/institutional problems of the state. Baum's account of this complex issue is quite good on the whole, although he often digresses, gets tangled up, or simply sprawls (this middle section of the book is much too long and badly edited). Its biggest problem is his assignment of causal primacy to values (the result of following Parsons), which leaves his explanation too little grounded in identifiable historical events such as the German reaction to the French Revolution and Napoleon's invasions, Germany's own failed revolutions of 1848 and 1918, and the social cleavages and political alignments which attended industrialization and national unification. Almost all his elements are assembled, but they do

not work as an explanation because they are out of order. His chapter on value dissensus contains a very interesting quantitative analysis of values manifested in popular novels—I hope something comparative will emerge soon from this direction—which is, unfortunately, marred because his "values" are so abstract (Parsons again) that his coders could not agree on them. Also, considering his thesis, an analysis of biographies of elites might have been more appropriate.

Baum concludes his study with some very suggestive but rather inconclusive considerations of a social theory of moral indifference as a phenomenon specific to modern society. Much of this discussion derives fairly straightforwardly from theories of mass society, but I was quite taken with his ideas about the problems of "managing trust" among "moral strangers" through the "symbolic media" of money, power, influence, and value commitments (pp. 299, 301, 306–14). I liked particularly his application of the notions of extension of credit and inflation, familiar in thinking of money, to the last three symbolic media. Trust (extension of credit) is necessary in a society too large for people to know each other, and when this trust is broken because relations become devalued (inflated), human strangers can themselves become devalued in the development of moral indifference, and a crisis becomes possible.

Baum's study represents a good step toward an adequate sociological account of the causes of the Holocaust, especially his attempt to combine theories of status anxiety with those of social and political development (although, as I say, I think he put them in the wrong order). And while it is not well-written or well-edited, it contains a wealth of material and lends itself well to further development along the same lines. Perhaps the main element still to be incorporated into an account of the Holocaust, after Baum, is not what failed to prevent it but what caused it.

Who Voted for Hitler? By Richard F. Hamilton. Princeton, N.J.: Princeton University Press, 1982. Pp. xv + 664. $50.00 (cloth); $16.50 (paper).

Michael Rogin
University of California, Berkeley

Nazism, as every social scientist knows, was a movement of the lower-middle class. Economically fragile and psychologically anxious, so the argument goes, members of the petite bourgeoisie felt trapped between large capital and the organized working class. Isolated and alone, they had lost their moorings in traditional culture and were in danger of losing their middle-class status as well. The lower-middle class, we are told, therefore joined mass, authoritarian, populist movements throughout Europe and America. In Germany it voted for Hitler.

Richard Hamilton refutes that view and challenges much other conventional social science wisdom in the process. He offers an interpretation of the rise of Nazism based on economic fears, political choices, and

organizational power rather than on social structural strains and psychological anxiety. *Who Voted for Hitler?* is a brilliant, important book that has much to teach us about the rise of Nazism in particular and about the loose links between social structure and political behavior more generally.

Hamilton first analyzes electoral data for 14 German cities. In 11 of these, including Berlin and Hamburg, the NSDAP (National Socialist German Workers Party) was strongest in the wealthiest neighborhoods and weakest among the working class. The lower-middle-class hypothesis would have predicted a bell-shaped curve, with Nazi support lowest in richest and working-class areas and strongest in the neighborhoods in between. That is not true for any city. It proved impossible to locate distinctively lower-middle-class voting units, and Hamilton is sensitive to the dangers of the ecological fallacy. But there is simply no evidence for a distinctively lower-middle-class Nazi vote. Cologne and Munich, the only cities where the wealthiest Protestants were not distinctively pro-Nazi, contained strongly anti-Nazi newspapers read by the upper-middle class. The respectable press in other cities, Hamilton shows, legitimized the NSDAP as an answer to the Communist menace and thus helped undermine the parties the papers themselves supported.

The precipitous decline of these so-called middle-class parties has been taken to demonstrate a petit bourgeois shift to the NSDAP. But these parties, Hamilton suggests, had mobilized voters across the class spectrum. One, the German People's party, was the traditional home of big business and the wealthy professions. The "middle-class" parties received significant support from the German equivalents of Tory workers, who then provided Hitler with substantial electoral backing.

The NSDAP received its greatest support, however, in the Protestant countryside. Do the votes of these independent farmers sustain the lower-middle-class hypothesis? Not according to Hamilton. First, rural Protestant areas moved en masse into the NSDAP in 1930. Farmers were characterized by community consensus rather than class division. Second, the hypothesis pictures· a deracinated, isolated urban petit bourgeois, who bears little resemblance to the traditional German small farmer. Third, the farmers were responding to an economic crisis, not status anxiety. And, finally, the NSDAP failed to mobilize Catholic farmers. One could, following Durkheim, save the hypothesis for the countryside by suggesting that the Catholic church integrated its communities, as Protestantism did not, and that Catholic farmers therefore suffered less anomie. Hamilton ignores that possibility in favor of a political explanation: he points to early Nazi anti-Catholicism and compares the Catholic church's opposition to the NSDAP with Protestant clergy's support for it.

Hamilton explains the rise of Nazism by political choices, not social psychological strains. He stresses the bourgeoisie's fear of the left, the impact of the depression, and the legitimation of the NSDAP by newspapers, local notables, and other elites. But he rejects decisively the view

that the Nazis were a creature of big business. The divided bourgeoisie lacked political hegemony; the Nazis, by contrast, had impressive, independent power. Hamilton's analysis of the Nazi movement again emphasizes political organizational over social psychological factors. That analysis forms, with the interpretation of electoral data, the second major contribution of this book.

Unlike the Communists, the NSDAP had a large, stable, committed cadre. The party trained its militants who, fanning out across Germany, made the NSDAP an organized presence comparable only with the (initially much more numerous) Social Democrats. The extraordinary organizational commitment of NSDAP members, according to Hamilton, is explained not by the apolitical abstractions of mass theory but by the impact of World War I. That impact began in war itself, was extended by the myth that betrayal on the home front deprived Germany of victory, and was brought home by the particular experience of returning soldiers themselves—first their physical humiliation at the hands of revolutionaries in 1918, and then their prolonged *Freikorps* (volunteer, government-sponsored military units) experience (until 1923) fighting domestic radicals throughout Germany and foreign enemies in the Ruhr. War, destruction, and the menace on the left became a permanent way of life for these ex-soldiers. The veterans initiated younger militants into that way of life in the party. The Nazi organizational weapon penetrated not only the towns but also the countryside (as Socialists and Communists did not), so that while Scandinavian farmers were voting Socialist, their German counterparts supported the NSDAP.

A major consequence of Hamilton's approach, though he acknowledges it explicitly only in a long footnote, is to downplay the importance of anti-Semitism among Nazi supporters. Only Frankfurt, with a tradition of anti-Semitism, lends possible support to the picture of a Jew-scapegoating petite bourgeoisie. Even there Hamilton prefers an economic explanation: competition with an atypically large Jewish lower-middle class. Anti-Semitism also played a role in the countryside, which Hamilton attributes to the alleged exploitation of farmers by Jewish middlemen. A non–economically based anti-Semitism contributed significantly to Nazi support only among (upper-class) university students. For the bulk of Nazi militants and voters, in the towns and cities, Marxism was a more important enemy than were the Jews. The Nazis created anti-Semitism, in Hamilton's words, rather than anti-Semitism creating the Nazis.

This book concludes by reflecting on the German exiles, liberal and Marxist, who blamed the lower-middle class for Nazism even before Hitler came to power. Such figures (some but not all associated with the Frankfurt School) needed to believe in a stable, liberal, upper class, just as their American counterparts did in the 1950s and 1960s. These German exiles and American scholars are Hamilton's targets, and he sometimes seems more indignant at their scapegoating of the lower-middle class than he does at Nazi scapegoating of the Jews. The petit bourgeois theory surely was a rescue operation which, by locating authoritarian predis-

positions among the lower-middle class, preserved both liberal and Marxist visions of the world. But if class, politics, and organization explain the triumph of the NSDAP, they hardly suffice to explain its character. There is a disjunction here (partly met by the discussion of war) between rational cause and insane consequence. If Hamilton is right to downplay anti-Semitism in the Nazi rise to power, for example, how would he analyze its genocidal triumph under Nazi rule? Mass theory might have its uses, were it detached from its special hostility to the lower-middle class and applied more generally to the NSDAP cadre, its leadership, and to the acquiescence in Nazi destructiveness by the society as a whole.

Peace against War: The Ecology of International Violence. By Francis A. Beer. San Francisco: W. H. Freeman & Co., 1981. Pp. xx + 447. $19.95 (cloth); $9.95 (paper).

Elise Boulding
Dartmouth College

In this valuable contribution to the field of conflict and peace studies, Francis A. Beer offers a historical epidemiology of peace and war. War is treated as a disease, a social virus, identifiable and treatable through the social science disciplines. At least, that is the optimistic note on which *Peace against War* begins. Subsequently, Beer develops an ecological model, postulating interactive relationships among aggregation, polarization, and militarization. Since these interactions seem to lead far more persistently to war than to peace, the possible intervention points leading to stable periods of peace appear few and meager indeed. However, the ecological analysis of conflict processes which Beer offers is an important new approach. The metaphor he uses for the social environment unnecessarily belies the initial ecological concept: He sees the social environment as "an immense machine—a world machine. Each of us, whether we struggle against it or not, is a part of it." This mechanistic conception of social technology, about which Beer is himself clearly ambivalent, is partly responsible for the gloomy optimism with which Beer views the prospects for controlling war in the future.

Nevertheless, Beer has made a profoundly important contribution to the study of global conflict processes. In addition to the ecological concept itself, his analysis of aggregation, the development of "larger, relatively integrated units out of smaller elements," is fresh and new. He shows how the integrative aspects of aggregation are countered by the differentiating aspects of the same process which involve inequality and instablity, creating cleavages in the larger system which lead to a condition of social polarization. That condition of polarization justifies the development of military elites and strong central control at the national level, alliances and war at the international level. Scholars who use the functionalist and integrative models of world order do not adequately take account of the

757

double-edged character of the differentiation that accompanies systems growth.

A second important contribution Beer has made to the study of global conflict processes is the bringing together of a variety of quantitative and semiquantitative indicators of ecosystem states and trends, some macrohistorical, some dealing with recent times. Teachers of global-systems courses will find this a splendid collection of indicators. (Data derivations are not adequately explained in the text for some of the trend charts, so teachers must be prepared to provide their own explanations for students.) In the 1940s, when the pre–World War I movement to develop indicators to monitor the health of the international system began receiving attention from peace researchers, it was thought that the development of an early warning system would be a major step toward a more peaceful world order. By comparing trend charts for different periods and different indicators, Beer shows how overoptimistic the monitoring hope was. Sometimes arms races lead to war, sometimes they do not. Sometimes rising GNP makes a nation more peaceful, sometimes not. Quantitative data continue to be important, but more sophisticated cross-impact analysis will be needed before they can be used for prediction.

A third contribution of *Peace against War* is a fine 60-page bibliography. Inevitably, because of the nature of publication timetables, more recent (post-1977) books are underrepresented. Pre-1977, however, the selection is excellent. The *Bibliography on World Conflict and Peace* (compiled by Elise Boulding, J. Robert Passmore, and Robert Scott Gassler [Boulder, Colo.: Westview, 1979]), which was originally a project of the ASA Section on the Sociology of World Conflicts, was being developed during the same time period at the University of Colorado and benefited at that time from Beer's advice.

To return to the book's "gloomy optimism": Beer discusses prospects for changes in the war system in a sober closing chapter on choices. As he points out, the range of human choice is restricted, and action must be undertaken in the face of great uncertainty. "We must live with the contradictory dynamics of our ecology and of our own choices and actions" (p. 325). While it is true that grounds for optimism are sparse, the apparent difficulty of breaking the negative aggregation-polarization-militarization cycle is partly owing to the fact that socialization and learning processes do not easily fit into an ecological mode. Beer draws with considerable competence on all the social science disciplines, but the choice of materials for inclusion from other disciplines is determined inevitably by his ecosystem framework. Over the past 15 years, a substantial literature has developed on learning processes and what is called somewhat simplistically "education for disarmament." This new field in fact addresses the problem of developing a more sophisticated understanding of how changes in mental imagery concerning war and peace come about. The underlying hypothesis of this field is that institutional changes are useless unless accompanied by changes in social imagery and accompanying behavioral predispositions. Beer touches on this in his discussion of exchange and

language but does not develop it fully. One of the several readers on disarmament education now available might be a useful supplement to the Beer book for teachers who wish to bring out this dimension.

Overall, this book is a major addition to the growing body of literature on peace and war processes. I recommend it highly for both research and teaching purposes.

Civil-Military Relations: Regional Perspectives. Edited by Morris Janowitz. Beverly Hills, Calif.: Sage Publications, 1981. Pp. 288. $25.00.

Michel L. Martin
Institut d'Etudes Politiques de Toulouse and *University of Chicago*

If it is now admitted that there could be no relevant scholarship in the social sciences without a comparative perspective, some areas still are little affected by this concern. In the growing field of the *res militaris,* a comparative effort has been attempted in the pioneering writings of men like Hans Speier, Alfred Vagts, Harold Lasswell, and their disciples, one of whom is Morris Janowitz. Yet these remain rare and, on the whole, the so-called comparative analyses of the military "are no more than a juxtaposition of various case studies on an ad hoc basis," to borrow a phrase from the introduction of *Civil-Military Relations: Regional Perspectives,* which gives an apt view of the "natural history" of the study of this important issue. Edited by Morris Janowitz, whose earlier *The Military in the Political Development of New Nations: An Essay in Comparative Analysis* (Chicago: University of Chicago Press, 1964) exemplifies his ongoing concern for a comparative view, the present work is precisely an attempt to transcend that flaw. It is a collection of essays by area specialists which seeks to offer the reader a comparative assessment of world patterns of civil-military relations, their nature and dynamics. Breaking with the type of transnational approach which heretofore has dominated the field and which consisted only of discriminating among the three worlds—Western, Eastern, and the developing one, the book uses a regional approach (a region being loosely defined as a geographical area with a minimal intracultural homogeneity). Such a level of analysis allows for distinguishing the most relevant specificities and uniformities from a macroanalytical standpoint while lending itself also to useful cross-regional as well as intraregional comparisons. The originality of this perspective lies in the fact that it emphasizes the cultural factors and their impact on the diversifying styles of civil-military relations; in that it amplifies Samuel Finer's *Man on Horseback* (London: Pall Mall, 1962), one of the first attempts to categorize such types of relations according to levels of cultural development.

The emerging picture is a complex one—which can, unfortunately, be rendered only sketchily here—especially in the nonindustrialized and industrialized worlds. In this regard, aside from the article on Africa by

David Laitin and Drew Harker which deals only with a particular aspect of the issue at hand, most of the papers concentrate on the Asiatic continent, both the Far East and the Middle East (this last also contains some remarks on North Africa).

Three patterns of civil-military interactions emerge in the Far Eastern area. For the northeast and southeast the examples Eugene Kim discusses are South Korea, Taiwan, Indonesia, and Thailand where the military generally dominates politics (under a form known as quasi-civilianized direct military rule) and which have been able to generate a stable, though limited politicized and participationist regime (a tutelary democracy, to use Edward Shils's wording) as well as a prosperous industrializing economy operating in a free-market framework. South Korea is prototypical in this respect, while Thailand is a rather deviant case.

Differing from these societies, which are fairly homogeneous and not incompatible culturally with elements of authoritarianism, are Thailand and other countries in South Asia: India, Pakistan, Sri Lanka, Indonesia, and Burma, studied by Ann Gregory and DeWitt Ellinwood in a thoroughly researched paper. Given the ethnic heterogeneity of their social fabric, these states are characterized by a low level of civic culture and polarized pluralism, which complicates, if not precludes, the search for legitimacy along with political order and government effectiveness. Although each state had its own history of political change, on the whole there was, over time, a shift from legitimacy-oriented civilian rule based on ethnic representation to order-minded military rule less concerned with the issue of representation (even remaining civilian regimes have tended to follow this trend); yet the latter has not been any more—actually rather less—successful than the former in setting up the conditions for economic development. In these countries, then, and particularly under praetorian rule, ethnic bargaining is the key prerequisite for the formation of a stable polity and growing economy.

The history of modern civil-military relations in China, the third Asian subregional model, is that of increasing differentiation, which for reasons linked to the political-military fusion called for by the revolution had been minimal to start with. Harlan Jencks analyzes in detail the various aspects of this critical issue—critical because it contains the sources of possible future intramilitary and civil-military tensions. Differentiation and subsequent strains materialize with the military search for its professional autonomy, especially among the mounting generations which, unlike the older, regard the Maoist ideal of "an obedient and politically activist army" impossible and being simultaneously "red and expert" as unprofessional.

The Arabic region is another original case of political-military interactions, the particulars of which are addressed by Fuad Khuri. For cultural and historical reasons, the pattern of such relations in this area (which includes both the Middle East and North Africa) is characterized by a rather limited degree of differentiation (except, perhaps, with respect to the top of the military hierarchy), and the military remains highly sensitive to the peasant-oriented outlook of the parent society. Hence,

when assuming power, the military tends to operate within the framework of existing traditional loyalties and cleavages and, as such, reveals itself as a mediocre agent of social and economic change, aside from "monolithic projects" with no infrastructural impact. Except for affording channels of social mobility, the military not only duplicates civilian practices and social organization with its sharp elite-mass distinction—as evidenced by the officer-soldier status differences—it also does little to put its technological and organizational potentials and skills to societally relevant uses.

Of all regions the industrial Western and Eastern countries present a more homogeneous and stable model of civil-military relations. Paradoxically, both styles have a number of resemblances—limited but worth mentioning. In both regions, the civilians traditionally control the military, though the procedures of control are different: in the latter, for instance, it operates via so-called penetration and indoctrination techniques; the military, moreover, has to maintain an adequate operational level. Interestingly, in both cases, too, this pattern of civil-military relations has been undergoing tensions recently. The military is found increasingly to act as a pressure group. In the Western area, as aptly shown by Gwyn Harries-Jenkins, the military has to compete for resources with other sectors of the society, notably with a growing welfare sector (of which it has been the main victim, one should say) at a time when the operational requirements of the Western alliance are changing—a situation which tends to affect deeply the organizational format of the military. In the Eastern area, as observed by Ivan Volgyes, the military is acting not only out of corporate self-interest but also as the repository of nationalist feelings, which in turn affects, if not conflicts with, the "internal dimension" of its operational role and posture as the surrogate agent of the Soviet Union's goals.

These papers, needless to say, have a far richer content than can be described in a necessarily limited space. The regional analysis offers a very adaptable tool for the understanding of the logic of political-military relations at a macroanalytical level. Yet the regionalist format, as defined in this book, could in some cases be reshaped to be made amenable to greater discrimination between patterns of civil-military relations which have more differences than similarities. North Africa, thus, could perhaps be separated from the Middle East, for beyond the commonality of religion and language, there are other relevant factors at work which differentiate the two regions. Africa may be another case. Though limited to a particular aspect of military rule (namely, the handling of secession), Laitin and Harker's paper touches on two opposite styles of rule which suggest two subregional models of civil-military interactions; Africa seems to shelter a rich variety of those. Latin America (which has not been included here although it was included in the conference where this book originated) is also an area with at least two fairly differentiated intraregional models (South and Center).

These are minor points. On the whole, Janowitz and his collaborators have prepared a very useful book that will afford not only the student of

761

armed forces and society but also those interested in comparative societal dynamics a set of working hypotheses with which to amplify their reflections on these matters.

Warriors to Managers: The French Military Establishment since 1945. By Michel L. Martin. Chapel Hill: University of North Carolina Press, 1981. Pp. xxii + 424. $29.00.

John H. Faris
Towson State University

By the summer of 1982, U.S. armed forces recruiting offices were virtually closing their doors to applicants who, a year or so earlier, would have been welcomed as recruits. At the same time, increasing numbers of service personnel were not permitted to reenlist, reflecting the oversupply of manpower. Yet in a recent volume produced by the American Assembly (*Military Service in the United States* [Englewood Cliffs, N.J.: Prentice-Hall, 1982, p. 195]), one of the most knowledgeable defense manpower experts, James Lacey, wrote that ". . . it is a near-certain prospect that the United States will resume a peacetime draft in the 1980s." Understanding why this might be so, and what alternatives might emerge, requires more than a calculation of demographic trends and unemployment rates. Sociologists and others concerned about public policy toward military manpower, conscription, and more generally about the relation of armed forces to society will find much value in Michel Martin's impressive study of institutional change in the French military since the end of World War II to the present.

The title, *Warriors to Managers,* is seriously misleading, for Martin's scope comprises much more than changes in the professional orientations of French officers. The thrust of Martin's thesis is the persistence of institutional forms in the context of fundamental institutional change, reflected primarily in the perpetuation of aspects of the "mass armed forces" institutional form, notably in the army, while other sectors, particularly in the navy and air force, moved toward a "force-in-being" pattern. This bifurcation was a consequence of a differentiation of missions (between maintaining a colonial empire and securing national defense) which resulted in uneven rates of technological change. The early 1960s stand as an important point of inflection in this process of change, as emphasis shifted from colonial *"grandeur"* to national security, and from a prenuclear ground-force orientation to a naval and air nuclear capability.

Martin's analysis is exemplary in both scope and detail. The discussion includes consideration of political, social, economic, and organizational factors, while organizing and presenting a diverse array of statistical data and facts. The comprehensiveness is manifested by the titles of the nine chapters—*"Grandeur* and Security," "Old and New Defense Roles," "Per-

sonnel, Arms, and Budget," "Manpower and Organization," "Citizen Participation in the Military," "Career Motivations and Conceptions," "Discipline and Authority," "Social Structure," and "Occupational Inheritance."

In his interesting conclusion, which considers future trends, Martin discusses the prospects of continuation of conscription. In the second half of the 1970s a significant process of "detabooization" of the draft had begun; that is, discussion of the possibility of ending the draft was no longer unheard. This, Martin argues, does not so much reflect a loss of legitimacy of conscription per se but, rather, an erosion of its classic attributes of universality and egalitarianism. Approximately one-third of young men were exempted, and these were disproportionately from families of the upper social strata.

This analysis is directly pertinent to contemporary issues in U.S. military manpower, where a similar, though reverse, process of "detabooization" is evident—a return to conscription has become thinkable, at least in part because of a parallel concern about social representativeness in the military. This suggests the hypothesis that a norm of civic obligation, which is as important for an all-volunteer force as it is fundamental to conscription, cannot endure (or re-emerge) if it is contradicted by a general perception of significant deviation from the associated principles of universality and egalitarianism. While this proposition leaves room for a wide variety of formulations of universality and egalitarianism, it has important implications for the questions of whether a return to the draft in the United States is likely, what sort of a system such a draft might be, and what the realistic alternatives (e.g., national service) are. Those interested in pursuing these questions will profit from reading Martin's account of the French case.

Why the Vietcong Fought: A Study of Motivation and Control in a Modern Army in Combat. By William D. Henderson. Contributions in Political Science, no. 3. Westport, Conn.: Greenwood Press, 1979. Pp. xxix + 163. $17.95.

Michael T. Anderson
U.S. Military Academy

Contending that there were few (if any) definitive research programs to investigate the motivation and cohesion of U.S. combat units and their importance in explaining the ultimate failure of the American army in Vietnam, William D. Henderson provides an interesting look into the organization and associated policies of the North Vietnamese People's Liberation Army (PLA) to explain why, in 1972, "one army endured and the other did not" (p. xv). *Why the Vietcong Fought* is a field study that describes the relationship between the PLA small combat units and the Communist party organization for which they fought, with a special focus

on the concepts of motivation and control. Crucial to this relationship was the role performance of the small unit cadres, who were judged to have effectively articulated the party goals and ideology with the goals of soldiers serving at the small combat unit level. This was accomplished primarily by making entrance into the primary group contingent on at least a manifest commitment to larger organizational values and norms and using particular leadership techniques to reinforce the congruence between primary and secondary group goals. What follows in Henderson's analysis is a description of the mechanisms and techniques used by the PLA to translate ideological party objectives into perceptions that would be meaningful to and capable of being internalized by the largely nonideo-logically motivated PLA rank-and-file soldier. This translation was effected by the carefully selected and well-trained cadre leadership, who performed the important requirements associated with the roles of "comradely leader" and "operator of the morale informant system."

As a result of the unique organization and associated policies of the PLA, Henderson concludes, the PLA was both combat effective and able to withstand the overwhelming combat power of U.S. forces in Vietnam. The PLA is credited as being a professional army "surpassing any other in its 'human' capabilities" (p. 119). Three strong forces are viewed as having shaped the motivation and control of PLA soldiers: the party organization, the promotion of primary-group solidarity at the small unit level, and the success achieved by the cadre in articulating the goals of the primary group with the strategic objectives of the party. The chapters in the book are organized around these major themes.

In addition to a review of hundreds of captured enemy documents relating to Vietcong activities, a principal source of data used to support the major hypotheses of the study comes from selected interviews of PLA prisoners of war and "ralliers" (defectors) conducted by the Rand Corporation between 1965 and 1967. The author adds his own observations from the perspective of a participant in the war.

This book deserves careful reading by anyone interested in a culturally informed perspective of organizational control, in particular, a military system of control in a combat environment. Henderson effectively extends the significant theme established more than three decades ago by Shils and Janowitz's "Cohesion and Disintegration in the Wehrmacht in World War II" (*Public Opinion Quarterly,* vol. 12 [1948]) of the importance of the articulation of primary- and secondary-group goals for combat effectiveness. Henderson has, in effect, replicated the findings of that study, extending its validity to another time and another place. Henderson's analysis also makes a significant contribution to the sociology of small-group behavior by focusing on the conditions and forms of leadership that facilitate the formation and preservation of primary-group relationships.

However, as a scientific analysis that attempts to attribute primary causation of ultimate Vietcong victory to the variables in the study (p. 121), the analysis has its share of flaws. To begin with, there is no attempt

made by the author to operationalize the dependent variable of "effectiveness" other than to equate it with the apparent ability of the PLA ultimately to endure in the face of overwhelmingly superior combat power. The conclusion that the PLA was, in fact, *combat* effective could have been strengthened had the author been able to demonstrate that the PLA was an effective fighting instrument during combat operations in addition to being able to survive as an organization. The reader is, instead, left to infer that endurance and combat effectiveness are one and the same. Second, even if one accepts the conclusion about the superior effectiveness of the PLA, a causal inference is made that the organization and policies of the PLA were most responsible for it. Given the author's source of data and research design, the less ambitious but more valid hypothesis could only be that the variables *may* have contributed to PLA survival.

A third, and perhaps most serious, limitation of Henderson's argument is the critical proposition that the combat motivation of the PLA soldiers was, in fact, oriented toward the achievement of PLA goals. The achievement of this alleged internalization is assumed to have resulted from the application of techniques and mechanisms of control inherent in the operating procedures of the PLA. The normative orientation of the participants thus becomes the criterion on which to base the success of socialization and resocialization efforts. Disappointingly, the author fails to support this central proposition and chooses instead to regard it as an assumption. In a sort of reverse logic Henderson argues, "Even though the PLA indoctrination and reeducation program was not always as successful as the party might have hoped, what is important is that the PLA soldier, whether he was a member of the in-group or the out-group, fought for PLA goals. This is supported by the fact that the PLA survived as an organization" (p. 117). Clear logic indicates that the answer to the question of whether the PLA was effective in achieving congruence of individual and organizational goals cannot simply be inferred from what one is trying to explain in the first place. Finally, there is the methodological problem of using data collected at an earlier period (1965–67) to explain a result measured at a later date (1972). That this is a valid concern is evidenced by the author's own conclusion about the deleterious effects of stress and hardship on the self-perceptions and cohesion of PLA units (p. 117). The burden of proof rests with the author in demonstrating that the bonds between the Communist party and the combat units withstood the escalation of the war effort between 1967 and 1972 effectively, that is, that these bonds were still intact at the time the conclusion is made. One might note that the U.S. army, during the same period the data were being collected for this study, was also described as possessing a high degree of cohesion and morale (p. xvi).

In spite of the reservations I have expressed, *Why the Vietcong Fought*, taken for what it is—a field study—is an important addition to analyses of the interrelationship between systems of motivation and control within a combat setting. Readers familiar with the research tradition illustrating the importance of primary-group cohesion and the nature of primary-

group goals for combat effectiveness will view the Henderson study as a major contribution to that tradition.

Ethnic Soldiers: State Security in Divided Societies. By Cynthia H. Enloe. Athens, Ga.: University of Georgia Press, 1980. Pp. xxi + 276. $19.00.

Douglas Lee Eckberg
University of Tulsa

Ethnic Soldiers is a first-rate study of the relationships between ethnicity and important aspects of the state. In it, Cynthia Enloe challenges the position that the state, through the military, serves as a primary agent of modernization, one that causes ethnic boundaries to break down and individuals' loyalties and identifications to become attached to the central nation-state. Quite the contrary, she makes a case for ethnicity's existence as an ongoing phenomenon that permeates diverse areas of decision making in states by elites. She comes close to proving her contention.

Enloe is especially interested in the relationships between the military and ethnic groups for two reasons: it is the military that is most clearly responsible for state security, and the military (at least in more modern states) is often portrayed either as ethnically neutral or as the great destroyer of ethnic identification. Her case for the military-ethnicity relationship is quite different. She contends that the perception of a given state's elite of the "trustworthiness" of various ethnic groups (i.e., where groups fall on an elite's "ethnic state security map"), the state's need for soldiers in extreme times, the need to consolidate consensus or fragment an opposition, and so forth, all help determine the relationship between ethnic groups and the military in various societies. Depending on the situation, elites may seek to develop or diminish ethnic identification, to play off groups against one another, or to convince peoples of their common nation-state citizenship.

In one example, Enloe states that, in premodern, colonial situations, it was common for elites to form ethnically homogeneous military units, often with members of relatively small ethnic groups, or geographically isolated ethnic groups, from a conquered population. These units were then used to control geographic areas distant from that group's original home. She supports this contention with evidence from the histories of such "martial races" as the Irish, the Kurds, the Cossacks, and the Gurkhas. However, in modern states there are growing pressures to develop ethnically heterogeneous militaries, a situation of some concern for elites since they do not wish to grant military training or weapons to "untrustworthy" or "incompetent" groups. In fact, domination of militaries (especially officer ranks) by members of elites or "safe" ethnic groups, continues to be a source of conflict in many modern states. Hence, ethnicity acts

differently here than in the situation above, but it remains a major concern for elites.

This should not be taken to indicate that Enloe sees ethnicity as merely a phenomenon to be manipulated by elites. Elites do manipulate ethnicity, but to some extent ethnic identification acts as a limiting factor in state security planning; it must be planned around. Also, mobilized ethnic collectivities attempt to maximize their situations within the state by selective affiliation with militaries.

The book is an excellent contribution to the comparative literature on ethnicity, with special significance for the literature dealing with the institutional context of discriminatory actions. It is well-organized, clearly written, and wonderfully cross-national and cross-period in scope. Moreover, it is copiously footnoted.

To be sure, the book has some weaknesses. It relies heavily on demonstration by example, which is a weakness since any counterinstances may be neglected. The emphasis on closed-elite control of events is not demonstrated fully; sometimes the elites in question go unnamed, and it is not always certain *who* has control of the organization of events. That is, Enloe does not challenge pluralist/compromise views of policy determination squarely. Moreover, some of her presentation amounts largely to the categorization, rather than the explanation, of events. And Enloe has a tendency to describe working hypotheses in terms that make them seem to be generally accepted laws. These are minor problems. Of more importance to me is the fact that, although she pretty clearly sees the "ethnicity" perspective as challenging the "modernization" perspective, Enloe never really concerns herself with the latter. That is, it might very well be that both sorts of processes are at work, but Enloe does not deal with this possibility, and it is not clear whether she would accept such a hypothesis. Certainly, her choice of evidence indicates, at the very least, a long-lasting role for ethnicity in issues of state security.

Praise the Bridge That Carries You Over: The Life of Joseph Sutton. By Shepard Krech III. Cambridge, Mass.: Schenkman Publishing Co., 1981. Pp. xxvii + 209. $8.95 (paper).

Margaret L. Andersen
University of Delaware, Newark

Praise the Bridge That Carries You Over is the history of a black man, Joseph Sutton (1885–1980), whose life spanned nearly a century and who lived in some of the rural all-black hamlets on the Eastern Shore of Maryland. Shepard Krech, the author, based the book on over 80 hours of interviews, conducted over a period of three years (1976–78), with Joseph Sutton. The bulk of the book consists of Sutton's edited narrative, but, together with the author's introduction and two appendixes (described below), the book provides a rich account of the formation and

social structure of rural black communities, the ethics and beliefs of their people, family networks, work, and recreational activities.

As a life history, the book is valuable because it provides an account of race relations from the perspective of the oppressed. And, because of the special quality of life histories, this account conveys the full range of emotions and thoughts that Sutton's experience encompasses. Thus, reading the narrative engrosses you in Sutton's wit, sensitivity, sharp insights, and humane world view. In addition, the two appendixes make the book especially valuable for scholars.

The first appendix is a historical overview of the development of racism on Maryland's Eastern Shore and in Talbot County, in particular. It details the political, economic, and social transformations that fostered racism in this region. In colonial Maryland, tobacco quickly emerged as the primary crop and was significant in creating a labor-intensive economy for which the practice of indentured servitude could not provide an adequate work force. The development of a plantation economy in this region emerged alongside a large free black population; for this reason, Maryland is a rich region for sociological study because it encompasses the social structural features of both a slave and a free society. Krech also details the emergence of the postslavery tenant farming system and its consequences for black labor; also, he reviews the racial violence and interracial tensions that predominated in the early 20th century in this area and throughout the South. Thus, along with Sutton's narrative, readers are given the full historical context for Sutton's experience.

In the second appendix, Krech carefully describes his research methodology. He includes information on his relationship with Sutton and the social context of the interviews, and he also verifies, through the use of archival records, many of the events and geneological relationships described in Sutton's narrative. This appendix also includes a discussion of Sutton's patterns of speech, their regional origins, and Krech's transcription of his language. Thus, the book is an exemplary model for the development of life history as a sociological method.

The book is an excellent resource for courses in race relations, sociology, and history, the individual in society, social change, and sociological research methods. Students who read it are moved by Sutton's experience, and it provides them with a context in which to examine their own lives and the history of racial privilege and oppression. By studying Sutton's life, they are better able to discard such stereotypes about black Americans as those which define them as not holding a work ethic, as living in unstable families, and as not forming strong community attachments. Sutton's life depicts vividly the history of racial segregation, violence, and injustice—historical facts too often unappreciated by today's white, middle-class students. For black students, Sutton's life history is an excellent case study of an experience that may be quite different from their own yet is still revealing of the collective experience of black Americans.

Joseph Sutton was born in 1885 in Copperville, a small and isolated

black hamlet in Talbot County on Maryland's Eastern Shore. Readers may recognize Talbot County as the home of Frederick Douglass who, as a young boy, was a slave on the plantation of Edward Lloyd—the largest slaveholder on the Eastern Shore. Sutton's father also worked as a slave for Lloyd, and Sutton's narrative recalls his father's and his elders' descriptions of slavery. Through the narrative we learn about the relations between slaves and free blacks. (Of the slaveholding states, Maryland had the largest proportion of free blacks.) Sutton graphically describes the slaves' quarters, slave labor in the fields and in white and black households, the physical brutality of slavery, slaves' fears of being sold further south, the influence of Quakerism on the slavery debate in Maryland, and the recruitment of slaves by the Union army. Sutton's discussion of slavery, based on the oral accounts of his community, shows us slavery from the slaves' perspective, a view that is a rare and yet necessary part of the historical record.

Following Emancipation, the black men and women who settled Copperville (and other black hamlets like it, still found throughout the Eastern Shore) continued to work as laborers on Lloyd's plantation and other nearby farms. Sutton himself, who began work at the age of nine, worked in several capacities, including oyster culling, carpentry, bridge construction, farming, chauffeuring, dog raising, and boating. A recurring theme throughout Sutton's narrative is the instability of work, dependent as it was on seasonal fluctuations in agricultural labor, periods of economic depression, and the individual whims of white employers—who often did not pay wages that were due and who always paid substandard amounts.

Despite these conditions, Sutton's community was a cohesive one where resourceful persons mustered whatever resources they could, including a strong work ethic and an attempt to become as self-sufficient as possible. Women's labor in this context is clearly central to the economic structure of the community, where they worked as midwives, hired domestics, and agricultural workers. However, the scarcity of jobs necessitated ongoing migration to the cities. Individual family members would leave, even if temporarily, to acquire better-paying jobs in the industrialized economy and would send money to their home communities (a pattern continuing to this day). As a result, households were fluctuating systems of social organization. As other research on the urban poor has shown (Carol Stack, *All Our Kin* [New York: Harper & Row, 1974]), household organization is shaped by economic conditions but is a resilient and elastic system of economic and social cooperation among individuals within the extended kinship network.

Other sections of Sutton's narrative discuss the racial antagonisms and extralegal forms of social control (lynching, terrorism, and violence) that characterized interracial relations in the early 20th century. Sutton renders a clear and chilling account of the dual system of justice that brought harsh sanctions against blacks alleged to have committed crimes, while pardoning acts of overt violence committed by whites against blacks.

Sutton's own philosophy of justice reverberates through these accounts, as his own world view is centered on the ethic of egalitarian justice. He says,

> They have it just as bad here as it is down south. . . . Just as bad. That prejudice, that's the stuff. . . . It don't pay to talk about prejudice too much. I don't guess I spoke of it half a dozen times. But I haven't forgot it, though. I still think of it. It don't pay to talk too much cause I could get somethin unjust just as well as the other fella. But, as I say, we got somethin comin, we ain't going to miss it. I may be gone, I hope I will be, but we got somethin comin. . . . They speak of justice. We got somethin comin and I hope I'll live to a ripe old age and be gone. That stuff can't go on all the time. There'll be an end to it someday. [Pp. 125–26]

Sutton never lived to see the just world he envisioned. He died in 1980 at the age of 95 and is buried among his ancestors, kin, and neighbors in a small graveyard beside Saint Stevens African Methodist Episcopal Church in Unionville, Maryland. Near his grave is another marker which now reads only, "Charlotte, aged 106 years." One wishes we had her story, too.

The Jack-Roller at Seventy: A Fifty-Year Follow-Up. By the Jack-Roller and Jon Snodgrass, Gilbert Geis, James F. Short, Jr., and Solomon Kobrin. Lexington, Mass.: Lexington Press, 1982. Pp. ix + 173. $17.95.

Oral History and Delinquency: The Rhetoric of Criminology. By James Bennett. Chicago: University of Chicago Press, 1981. Pp. xv + 363. $28.50.

Jim Thomas
Northern Illinois University

It is not without some justification that the method of oral history has remained little used by social scientists. As Gilbert Geis observes in a concluding commentary in *The Jack-Roller at Seventy* (p. 132), even the rich material collected by Clifford Shaw wore thin in that "eighty-five cases histories are said to remain in the Shaw-McKay archives, and there is no clamor to see them into print because they offer very little that is new and much that is repetitive." Renewed interest in the Chicago School tradition of ethnographic research, however, has been accompanied by an awareness of the utility of oral histories, especially in criminology.

There remains little agreement on the domain and definition of criminological oral history, which James Bennett defines in *Oral History and Delinquency* as ". . . a statement about a life in which acts considered delinquent by the individual's surrounding society have a place, told by someone either under seventeen or usually not much older than that"; one who edits such statements is a "criminologist" (p. 5). Both of these

volumes exemplify how oral histories can be, or have been, used effectively by criminologists.

Bennett's work is the more ambitious of the two. What apparently began as an evaluation of Clifford Shaw's case histories of Chicago delinquents became an expanded study of the *history* of oral histories. Bennett, aware that a history reflects at least as much about the subject of knowledge as its object, identifies four methods, each embodying different research programs, by which an oral history might be written (rhetorical, logistic, generic, and dialectical). Although he attempts to integrate all four approaches, he acknowledges that his own stance is a "rhetorical method," which he follows with consistency and cogency. Although Bennett does not explicitly define "rhetoric," he implicitly employs the Aristotelian epideictic form, in which rhetoric refers to the expressive quality of language through which we display messages to an audience. Bennett's intent, although not developed adequately, is to demonstrate the symbolic power of language to generate moral and political meanings in ordinary discourse. Although Bennett's discussion of rhetoric and its power to shape meanings for a symbol-interpreting audience is minimal, his ability to illustrate the moral and political content embedded in the discourse of oral historians is impressive. Bennett's comprehensive description begins with the 19th-century street biographies of Henry Mayhew, John Clay's descriptions of convicts, and delinquents' "snitching bees" of early 20th-century judge Benjamin Lindsey. His chapter-length discussions of these early exemplars of oral history not only provide a thick description of the method but also display, as he promises in his introduction, the relationship between oral history as a mode of discourse and the axiological, ontological, and practical implications conveyed by the various attitudes the recorders bring to their objects.

Bennett does not draw from (or even mention) the works of linguists, semiologists, or rhetoricians, and this is, in my view, a weakness of his work. It may be unfair to criticize authors for not writing the book we would like to have read, but nonetheless, had he grounded his discussion even minimally in the framework of, for example, Kenneth Burke or similar writers, he could have strengthened his analysis dramatically by displaying tensions between the content of communication and the political, ideological, and organizational structures such messages proffer. Especially because Bennett's exemplars were reform oriented, an explicit grounding would have transformed this "history of histories" into a volume that provided social insights not only into social conditions and criminal justice policy but into changing forms of research discourse as well. As a consequence of this omission, Bennett's work fails to grasp the *dialectic* among rhetoric, self-display (both of the recorders and those recorded), and the social context in which such histories take on a social meaning for a social audience.

The volume by the Jack-Roller and Jon Snodgrass, although a life history by Bennett's definition, is considered a mere "autobiography" in

a commentary by James Short (p. 135), a point which Snodgrass acknowledges. Such semantic squabbling, however, seems irrelevant in the context of Bennett's observation that ". . . no one has the power or right to decree that a more consistent [terminological] usage is definitive unless it can be proved more productive in many fields than the existing chaos" (p. 5).

Stanley, the subject of this work, was the original Jack-Roller in Shaw's earlier study, published in 1930. Snodgrass met Stanley by chance nearly a half-century later and enticed him to participate in a lengthy series of interviews. Stanley was a teenage "jack-roller," that is, one who "mugged" drunks or other vulnerable victims. Snodgrass's history of Stanley, like Shaw's, provides a self-account of one delinquent's perception and cognition of existence, motivation, and justification. Although we are presented with those incidents which often make a 75-year life course seem more dramatically episodic than it may have actually been (for such episodes provide the organizing logic of this method), it is clear, nonetheless, that throughout his life he was institutionalized, victimized, abusive, and generally incapable of normal social intercourse. Stanley's later life, despite Shaw's claim that the young jack-roller had "gone straight," indicates that his earlier personality problems troubled him throughout his life, in that he was never able to avoid personal or social problems and live a routine life.

The additional commentaries by Gilbert Geis, James Short, and Solomon Kobrin provide an interesting, although occasionally somewhat gratuitous, conclusion to this work. Geis examines the impact of Shaw's work(s) and criticizes the method of oral history. From his cryptic conclusion that Snodgrass provides an "object lesson in some of the strengths and pitfalls" (p. 132), we must assume that he remains uncomfortable with the method. He suggests that Shaw's oral histories are overrated, and neither Shaw himself nor his epigones, he argues, seemed particularly influenced by the Stanley history. Short, although noting problems with the life-history approach, acknowledges, nevertheless, that life histories provide valuable additional data by which researchers can expand their knowledge. Kobrin, himself an earlier analyst of delinquency in Chicago, defends oral histories and illustrates how the approach can be employed to expand our understanding of, for example, labeling, social control, strain, and cultural-transmission theories. All three commentators focus their discussion more on Shaw and the life-history approach in general than on Snodgrass's work specifically.

One might fault Snodgrass for failing to develop Stanley's material more fully in a concluding analysis or for not bringing the analytic insights of his own earlier work on Chicago criminology to bear on his topic. Such material was, however (according to a personal communication from the author) excluded from the volume by the publisher (and was published in a special issue on Chicago sociology in *Urban Life* [January 1983]). Each of these two volumes, despite its flaws, reveals several reasons why oral histories remain an invaluable method. Each of these two vol-

umes, despite its flaws, is of exceptional utility for developing oral history methodology. First, at a minimum, oral histories provide useful supplementary data. The richness of the Stanley story, for example, is transformed into a longitudinal analysis by Snodgrass and provides researchers with a body of heuristic material. Furthermore, unlike the self-aggrandizing, cathartic, or exculpatory apologies of conventional autobiographies by other offenders (e.g., Jack Henry Abbott, Piri Thomas, Claude Brown, Eldridge Cleaver, Richard Nixon), oral histories provide a means of critique, clarification, elaboration, and verification. Second, as Howard Becker reminds us in his introduction to the 1966 edition of Shaw's *Jack-Roller*, "By truly entering into Stanley's life, we can begin to see what we take for granted (and ought not to) in designing our research—what kinds of assumptions about delinquents, slums, and Poles are embedded in the ways we set the questions we study." Oral histories thus sensitize researchers to methodological problems by providing knowledge of a subculture to which we bring a set of self-centric assumptions. Third, as Kobrin cogently argues in his contribution to the Snodgrass volume, oral histories furnish background information by which we may develop testable hypotheses or refine broader theoretical formulations. Fourth, oral histories can provide a type of data that seems to have gone out of style, namely, the construction of a detailed imagery of a specific scene to which most academics have little access. As Bennett argued, early oral histories, such as Mayhew's, were written largely for a lay audience and were intended as a means of bridging gaps between social groups. Finally, oral histories enable researchers to examine the accounts of persons whose behavior may require managing or justifying to outsiders.

Neither author even suggests that oral histories are more beneficial than, or should replace, other research techniques. Taken together, however, these two volumes illustrate that oral histories remain a powerful supplement to both inductive and deductive research, and both works are important references and models for further developing the oral history method.

NEW PRICE.—The price of the cloth edition of *Access to Power: Cross-national Studies of Women and Elites,* edited by Cynthia Fuchs Epstein and Rose Laub Coser (London: Allen & Unwin, 1981), reviewed in July 1983 (*AJS* 89:256–59), is no longer $27.50. It is now $12.50.

Choosing texts for next ·
Remember these
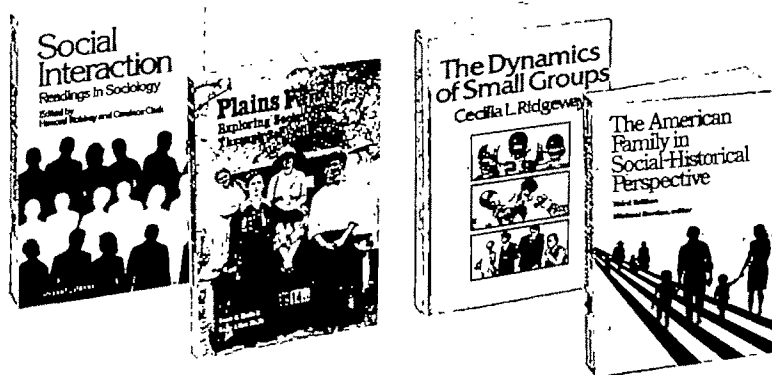
The American Family in Social-Historical Perspective
Third Edition

Edited by MICHAEL GORDON, University of Connecticut

Like its highly successful predecessors, the new edition of this anthology will help students understand how the contemporary American family has evolved—where there is continuity with the past as well as discontinuity, and why. The 25 selections included here represent the best articles on the subject by respected scholars in both history and sociology. Organized topically, the readings cover the life-span from childhood to death, including such key areas as youth, sex, marriage, divorce, and ethnic diversity. Approximately half the selections are new to this edition and represent the latest trends and scholarly developments in the field. An introductory essay by Professor Gordon surveys the current research in the field and its implications for understanding the American family.
Paperbound. 531 pages   1983

Plains Families
EXPLORING SOCIOLOGY THROUGH SOCIAL HISTORY

SCOTT G. McNALL and SALLY ALLEN McNALL, both of the University of Kansas

In this unique new textbook the authors draw on a wide range of historical documents of individual families—diaries, letters, personal records, and interviews—as a basis for the study of key sociological concepts and theories. The book is arranged in alternating historical and sociological chapters. The historical chapters present the authentic family documents; the sociological chapters then use these documents to illuminate such concepts as culture and social structure, socialization, stratification, sex roles, and social change. This unusual integration of sociology and social history brings to life abstract ideas, terms, and theories, and will enable students to understand and appreciate the relationship of individuals and family groups to society as a whole.
Paperbound. 328 pages   1983
Instructor's Manual available

semester?
from St. Martin's Press

The Dynamics of Small Groups
CECILIA L. RIDGEWAY,
University of Wisconsin—
Milwaukee

This important new text presents a comprehensive, eclectic analysis of small groups as vital units within the larger social structure. As a foundation, Professor Ridgeway provides a thorough and up to date examination of the theoretical approaches to small groups and the research methodology in the field. Using three main theoretical approaches—symbolic interactionism, exchange theory, and functionalism—to explain basic concepts throughout, she integrates the major studies and theories into a coherent analysis and interpretation of group behavior. The book deals extensively with primary groups as well as with task groups, and discusses group culture and the role groups play in society so that students can better understand the groups they must deal with in their own lives. Boxed inserts and numerous examples complement and illustrate the text, providing a concrete context for the abstract theoretical material.
402 pages   1983

Social Interaction
READINGS IN SOCIOLOGY
Second Edition

HOWARD ROBBOY,
Trenton State College

CANDACE CLARK,
Fairleigh Dickinson University

This anthology brings together 40 outstanding articles on major topics in sociology. Chosen for their · intrinsic interest and readability— and to provide a careful balance between the classical and contemporary—the readings will both serve as vivid illustrations of sociology at work and expose students to the intellectual excitement of the discipline. The book features a strong—though not exclusive—emphasis on symbolic interaction, thus providing students with a consistent approach to the various topics discussed in the articles. The editorial apparatus helps students integrate the readings with general concepts of sociology: there are review questions for each selection, suggested readings at the end of each part, and brief biographies of the contributing authors at the end of the book.
Paperbound. 597 pages   1983
Instructor's Manual available

A DIVERSITY FROM COLUMBIA
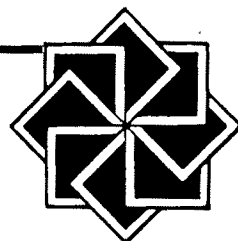
FROM THE SOCIOLOGY OF SYMBOLS TO THE SOCIOLOGY OF SIGNS
TOWARD A DIALECTICAL SOCIOLOGY

Ino Rossi. A major theoretical work which draws upon what the author identifies as the valid aspects of the structural and post-structural phases of semiotics in order to derive a "dialectical structuralism" as an alternative to some of the great classical sociological paradigms. 352 pp., *A King's Crown Paperback,* $15.00; $45.00 cl

LUKÁCS REAPPRAISED

Agnes Heller, Editor. A collection of essays written by distinguished Hungarian scholars who knew Lukács personally. "Anyone familiar with and interested in Lukács will find this volume of great interest and importance." — *Paul Breines.* 192 pp., $12.00 pa, $25.00 cl

SOCIETAL STRATIFICATION
A THEORETICAL ANALYSIS

Jonathan Turner. A study of the elements of social stratification, particularly the unequal distribution of money, wealth, power, and prestige, which identifies and defines the basic processes behind them and offers an investigation of the theoretical principles underlying those processes. 216 pp., *A King's Crown Paperback,* $13.50; $27.00 cl

WHITE RACISM
A PSYCHOHISTORY; SECOND EDITION

Joel Kovel. A new paperback edition of this important work, nominated for the National Book Award. Includes a major new chapter. *A Morningside Book.* 350 pp., $8.95 pa, $27.50 cl

MARXISM AND CLASS THEORY
A BOURGEOIS CRITIQUE

Frank Parkin. "Parkin has succeeded not only in offering us an up-to-date 'bourgeois' critique of Marxism but also in elaborating a genuinely 'neo-Weberian' position" — *Contemporary Sociology.* 217 pp., $10.00 pa, $24.00 cl

RELUCTANT REBELS
COMPARATIVE STUDIES OF REVOLUTION AND UNDERDEVELOPMENT

John Walton. Combines historical studies of national revolts in the Third World with a theoretical interpretation that demonstrates why these events belong to the proper study of revolution. 300 pp., $10.00 pa, $25.00 cl

ADOLESCENT SEX ROLES AND SOCIAL CHANGE

Lloyd B. Lueptow. An exhaustive study of adolescent sex roles based on a survey of some 5600 graduating high school seniors between the mid-1960s and mid-1970s. 352 pp., $35.00

GUESTS IN THE DRAGON
SOCIAL DEMOGRAPHY OF A CHINESE DISTRICT, 1895–1946

Burton Pasternak. Building on Arthur Wolf and C.S. Huang's, *Marriage and Adoption in China, 1845–1945,* this study focuses on the relationships among social institutions, domestic decision-making within the family, and demographic trends in rural Taiwan. 208 pp., $35.00

110 LIVINGSTON STREET REVISITED
DECENTRALIZATION IN ACTION

David Rogers and Norman H. Chung. Picking up where the senior author's *110 Livingston Street* (1968) left off, this new study offers a timely and reliable examination of the classic problems of bigness, overcentralized bureaucracy, and insulated professionalism that can be found in government and education nearly everywhere today. NEW YORK UNIVERSITY PRESS. 264 pp., $20.00

COLUMBIA UNIVERSITY PRESS
136 South Broadway, Irvington, NY 10533

2 1 AUG 1985

McGraw-Hill Sociology

A ROSTER OF FIRST-RATE TEXTS.

SOCIOLOGY, Sixth Edition
PAUL B. HORTON and **CHESTER L. HUNT.**
both Professors Emeritus, Western
Michigan University.

Now in its sixth edition, this leading introductory text presents a well-written, comprehensive overview of the major topics in sociology, and a descriptive treatment of research. Containing a wealth of new and updated information.

CRIMINOLOGY
ROBERT L. BONN.

CITIES AND SUBURBS
HARVEY M. CHOLDIN.

SOCIOLOGY
RICHARD T. SCHAEFER. Western Illinois
University.

This provides a balanced, comprehensive introduction to the discipline. A combination of features makes it a truly distinctive text: sound scholarship; current research; clear, lively writing; and wide integration of concepts and theories with vivid, contemporary examples. An extensive program of supplements supports the text.

EXPLAINING CRIME.
Third Edition
GWYNN NETTLER. Professor Emeritus, The
University of Alberta.

CHILD WELFARE: POLICIES AND PRACTICE, Third Edition
LELA B. COSTIN. The University of Illinois
School of Social Work; and **CHARLES A.
RAPP.** The University of Kansas School of
Social Welfare.